REPORT RESUMES

ED 020 498 48 AL 001 072

EXPERIMENTAL ANALYSIS OF THE CONTROL OF SPEECH PRODUCTION AND PERCEPTION. PROGRESS REPORT NO. I, FEBRUARY 1--SEPTEMBER 1, 1961.

BY- LANE, HARLAN AND OTHERS
MICHIGAN UNIV., ANN ARBOR,COLL.OF LIT.,SCI.,ARTS
REPORT NUMBER ORAP-04411-1-P PUB DATE SEP 61
CONTRACT OEC-SAE-9265
EDRS PRICE MF-$1.00 HC-$10.56 262P.

DESCRIPTORS- *LANGUAGE RESEARCH, *AUDITORY DISCRIMINATION, OPERANT CONDITIONING, *PROGRAMED INSTRUCTION, TEACHING MACHINES, APHASIA, *SECOND LANGUAGE LEARNING, LANGUAGE INSTRUCTION, STRESS VARIABLES, ACOUSTICS, ORAL COMMUNICATION, AUDITORY PERCEPTION, SPEECH, PSYCHOLINGUISTICS, EXPERIMENTAL PROGRAMS,

THIS DOCUMENT IS THE FIRST IN A SERIES REPORTING ON PROGRESS OF AN EXPERIMENTAL RESEARCH PROGRAM IN SPEECH CONTROL. THE TOPICS DISCUSSED ARE--(1) THE DISCONTINUITY OF AUDITORY DISCRIMINATION LEARNING IN HUMAN ADULTS, (2) DISCRIMINATIVE CONTROL OF CONCURRENT RESPONSES--THE RELATIONS AMONG RESPONSE FREQUENCY, LATENCY, AND TOPOGRAPHY IN AUDITORY GENERALIZATION, (3) EFFECTS OF CHANGING VOWEL PARAMETER ON PERCEIVED LOUDNESS AND STRESS, (4) OPERANT RECONDITIONING OF A CONSONANT DISCRIMINATION IN AN APHASIC, (5) TEACHING MACHINES AND PROGRAMMED LEARNING, (6) SOME DIFFERENCES BETWEEN FIRST AND SECOND LANGUAGE LEARNING, (7) TECHNIQUES OF OPERANT CONDITIONING APPLIED TO SECOND LANGUAGE LEARNING, (8) EXPERIMENTATION IN THE CLASSROOM--GUIDELINES AND SUGGESTED PROCEDURES FOR THE CLASSROOM TEACHER. (SEE RELATED DOCUMENTS ED 003 883 THROUGH ED 003 887 AND AL 001 073 FOR LATER PROGRESS REPORTS.) (DO)

(No Bureau number

04411-1-P

PA48

THE UNIVERSITY OF MICHIGAN

ED020498

COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS
DEPARTMENT OF PSYCHOLOGY
AND
COMMUNICATION SCIENCES LABORATORY

Progress Report I

February 1 - September 1, 1961

Experimental Analysis of the Control of Speech Production and Perception

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Under Contract with:

Language Development Section
U. S. Office of Education
Contract No. SAE-9265

Administered through:

September 1961

OFFICE OF RESEARCH ADMINISTRATION · ANN ARBOR

AL 001 072

THE UNIVERSITY OF MICHIGAN
COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS
Department of Psychology
and
Communication Sciences Laboratory

Progress Report No. I
February 1 - September 1, 1961

EXPERIMENTAL ANALYSIS OF THE CONTROL OF SPEECH PRODUCTION AND PERCEPTION

ORA Project 04411

under contract with:

LANGUAGE DEVELOPMENT SECTION
U.S. OFFICE OF EDUCATION
CONTRACT NO. SAE-9265

administered through:

OFFICE OF RESEARCH ADMINISTRATION ANN ARBOR

September 1961

CONTENTS

- STAFF
- RESEARCH COMPLETED
  - ON THE DISCONTINUITY OF AUDITORY DISCRIMINATION LEARNING IN HUMAN ADULTS
  - ON THE DISCRIMINATIVE CONTROL OF CONCURRENT RESPONSES: THE RELATIONS AMONG RESPONSE FREQUENCY, LATENCY, AND TOPOGRAPHY IN AUDITORY GENERALIZATION
  - THE EFFECTS OF CHANGING VOWEL PARAMETERS ON PERCEIVED LOUDNESS AND STRESS. I: DOES AUTOPHONIC LEVEL AFFECT THE LOUDNESS FUNCTION?
  - THE EFFECTS OF CHANGING VOWEL PARAMETERS ON PERCEIVED LOUDNESS AND STRESS. II: SOUND PRESSURE, SPECTRAL STRUCTURE, AND AUTOPHONIC LEVEL
  - THE EFFECTS OF CHANGING VOWEL PARAMETERS ON PERCEIVED LOUDNESS AND STRESS. III: VOCAL MATCHING OF STRESS PATTERNS
  - OPERANT RECONDITIONING OF A CONSONANT DISCRIMINATION IN AN APHASIC
  - TEACHING MACHINES AND PROGRAMMED LEARNING
  - SOME DIFFERENCES BETWEEN FIRST AND SECOND LANGUAGE LEARNING
  - TECHNIQUES OF OPERANT CONDITIONING APPLIED TO SECOND LANGUAGE LEARNING
  - EXPERIMENTATION IN THE LANGUAGE CLASSROOM: GUIDELINES AND SUGGESTED PROCEDURES FOR THE CLASSROOM TEACHER
- RESEARCH IN PROGRESS
- APPENDIX A. RESEARCH IN PROGRESS. INITIAL FINDINGS
  - SELF-SHAPING OF VOCAL BEHAVIOR
  - PROPERTIES OF THE VOCAL RESPONSE CORRELATED WITH RATE OF EMISSION
  - FOREIGN ACCENT AND SPEECH DISTORTION
- APPENDIX B. SPECIAL ACTIVITIES

STAFF

Staff

During the past seven months, the following have been members of the research and administrative staff of the project.

H. L. Lane, Ph.D., Director
D. V. Cross, M.A., Research Assistant
P. G. Shinkman, M. A., Research Assistant
B. A. Schneider, Technical Assistant
D. R. Brinkman, Technical Assistant
D. J. Moore, Technical Assistant
C. L. Mickelson, Secretary

The invaluable assistance of the staff and facilities of the Communication Sciences Laboratory (G. E. Peterson, Ph.D., Director) is gratefully acknowledged.

RESEARCH COMPLETED

ON THE DISCONTINUITY OF AUDITORY DISCRIMINATION LEARNING IN HUMAN ADULTS

ON THE DISCRIMINATIVE CONTROL OF CONCURRENT RESPONSES: THE RELATIONS AMONG RESPONSE FREQUENCY, LATENCY, AND TOPOGRAPHY IN AUDITORY GENERALIZATION

THE EFFECTS OF CHANGING VOWEL PARAMETERS ON PERCEIVED LOUDNESS AND STRESS. I: DOES AUTOPHONIC LEVEL AFFECT THE LOUDNESS FUNCTION?

THE EFFECTS OF CHANGING VOWEL PARAMETERS ON PERCEIVED LOUDNESS AND STRESS. II: SOUND PRESSURE, SPECTRAL STRUCTURE, AND AUTOPHONIC LEVEL

THE EFFECTS OF CHANGING VOWEL PARAMETERS ON PERCEIVED LOUDNESS AND STRESS. III: VOCAL MATCHING OF STRESS PATTERNS

OPERANT RECONDITIONING OF A CONSONANT DISCRIMINATION IN AN APHASIC

TEACHING MACHINES AND PROGRAMMED LEARNING

SOME DIFFERENCES BETWEEN FIRST AND SECOND LANGUAGE LEARNING

TECHNIQUES OF OPERANT CONDITIONING APPLIED TO SECOND LANGUAGE LEARNING

EXPERIMENTATION IN THE LANGUAGE CLASSROOM: GUIDELINES AND SUGGESTED PROCEDURES FOR THE CLASSROOM TEACHER

ON THE DISCONTINUITY OF AUDITORY DISCRIMINATION LEARNING IN HUMAN ADULTS

Harlan Lane
The University of Michigan

ON THE DISCONTINUITY OF AUDITORY DISCRIMINATION LEARNING IN HUMAN ADULTS

Harlan Lane
Communication Sciences Laboratory[1]
The University of Michigan

The truism that the adult human is an organism with an extremely complex history of discrimination training seems never to wear thin. This purely intraverbal sequence is evoked as often by the recalcitrant student as by the recalcitrant data point. Keller and Schoenfeld (1950) were presumably contending with both when they wrote: "In some instances [of verbal learning] the rate of improvement is so dramatic as to obscure the fact that essentially the same basic principles are involved in verbal as in nonverbal behavior." The same basic principles may apply to verbal and nonverbal, to human and infra-human learning alike, but in practice the differences between the two kinds of learning may play an important role. Thus, an increasing number of investigations in programmed instruction, where the verbal repertory of the subject is of central importance, suggest that human and infra-human learning are not at all points isomorphic. Wherein lies the difference may be a hard task for research but an understanding of the differences should make the underlying continuity of behavior more plausible.

Recent research in our laboratory on five diverse problems in verbal behavior has revealed some of the properties of discrimination learning in human subjects while under the influence of their extremely complex history. One generality that arises is the "dramatic rate of improvement" to which Keller and Schoenfeld refer. The present study describes this discontinuity in auditory discrimination learning and five conditions under which it occurred.

1. The Discrimination of Spanish Vowels

Six Spanish vowels (a, e, a, ae, i, a), rendered by a linguist, were recorded in irregular order at four-second intervals on a two-track tape recorder (Uher IIIA). One of the vowels, /a/, was designated as S^D; it appeared 30 times on the tape while each of the other vowels appeared six times. A four-second coding tone was recorded on the second track of the magnetic tape adjacent to each presentation of the S^D. During playback, the vowel signals were applied to a high-fidelity binaural headset worn by the subject, while the coding tone operated relay circuitry. The subject was seated in an anechoic chamber in front of a microphone and a counter. Each vocal response triggered a voice-operated relay; if the VOR operated while the coding tone was on, S received one point on his counter. Latencies were measured from the onset of the stimulus to the onset of the response with an accuracy of ±5 msec. (Hewlett-Packard 522B frequency counter). Three male undergraduates served individually in sessions lasting 30 minutes. They were instructed to respond, by saying /ka/, so as to accumulate points on the counter. The series of 60 stimuli was presented repeatedly until S made less than four errors (failed to respond to S^D or responded to S^Δ) in a given trial; the experiment was then terminated.

If the subject were learning to discriminate among the vowels of Spanish we might expect the two dependent variables, conditional probability of response and latency of response, to change during the course of conditioning in the following fashion. The probability of responding to S^D should be maintained or increase while the probability of responding to S^Δ should gradually decrease. "...extinction is the hallmark of discrimination—responding to S^Δ extinguishes while responding to S^D is maintained." (Keller and Schoenfeld, 1950). We may expect, as well, concomitant changes in response latency; those to S^D should decrease, those to S^Δ increase.

Figures 1, 2, and 3 show that our expectations are not borne out by the course of vowel discrimination learning with adult humans. Each figure plots, for one subject, the cumulative per cent of stimuli responded to in S^D and S^Δ as a function of trials, and the cumulative latency of consecutive responses in S^D and S^Δ. From the outset of the experiment, S_1 responded to S^D nearly 100 per cent of the time; responses to S^Δ fell abruptly after trial 1. There was no systematic change in the latency of responses to S^D or S^Δ. Subject 2 also showed nearly 100 per cent responses to S^D and an initial drop in responses to S^Δ after the first trial. S^D and S^Δ latencies do diverge during the first seven responses, but subsequently there are no S^Δ responses at all. Subject 3 presents the same picture of abrupt discrimination learning, although the change in S^D and S^Δ latencies during the first seven responses is more marked.

It was obvious to us that we were effecting a change in the discriminative behavior of our subjects, but the change took place so abruptly that it seemed appropriate to call it discrimination transfer than discrimination learning. Our natural inference was that our subjects had "an extremely complex history of discrimination training" and that, somewhere in their vast experience with speech signals, they had acquired the discriminative repertory we were now but sampling. We therefore sought an auditory continuum that did not play a major role in the discrimination of speech sounds, in the hope of observing human discrimination learning in a more pristine form. Vowel rise-time was our choice.

2. The Discrimination of Rise-time

The arrangement of apparatus, the format of the stimulus tape, the subjects and the instructions were exactly the same as in experiment 1; only the stimulus variable was changed. Six rise-times for the vowel phoneme /a/ were obtained by gating the S^D of experiment 1 with an electronic switch (Grason-Stadler 8295119) with variable rise-decay time. All six stimuli had the same duration (150 msec.) and the same amplitude (± 2 db). The rise-times of the gated signals were determined by processing the tape with an average speech power circuit (integrating time, 10 msec.), displaying the output voltage as a function of time on a calibrated oscillograph (Minneapolis-Honeywell Visicorder), and measuring the attack slope of each signal with a protractor. These slopes were then converted to db/msec. by means of the calibration. The values employed for the rise-time variable were: 0.8, 1.0, 1.3 (S^D), 1.6, 1.8, and 2.6 db/msec.

Figure 4 shows the per cent of responses to S^D and to S^Δ emitted by S_1 during his twelve trials. The subject responded about 65 per cent of the time to S^D and about 20 per cent of the time to S^Δ; these percentages did not change appreciably until the last trial when only 33 percent of the S^D's and

2 per cent of the S^{Δ}'s were responded to. Figure 5 gives the mean latencies on the twelve trials for S_1. It will be seen that S^D and S^{Δ} latencies do not diverge after trial 2 but vary jointly as a function of some third, uncontrolled variable. A breakdown of the latencies of responses to S^D and S^{Δ} is given in Fig. 6 for the first few trials. Far from increasing, S^{Δ} latencies decrease during trial 1, fluctuating slightly thereafter. S^D latencies show an initial decrease but do not change appreciably over the rest of the session. If extinction is the hallmark of discrimination, then S_1 did not learn a discrimination at all.

The findings for S_2 and S_3 are similar. However, one change in procedure was introduced with these two subjects. On trial 7 for S_2 and on trial 4 for S_3, the subject was informed that he could now obtain several points during the interval following each S^D by responding rapidly at that time. Figure 7 shows the effect of this instruction on the per cent of stimuli responded to by S_2. Prior to trial 7, S_2 had been responding to 90 per cent of the S^D's and 50 per cent of the S^{Δ}'s; subsequently, he responded to all stimuli. The latency-of-response functions for S_2 prior to trial 7 are not shown since they resemble closely those for S_1. It was not possible to measure latency following trial 7, since the subjects' runs of responses following S^D did not terminate until after the onset of the next stimulus.

Figure 8 shows, for S_3, the per cent of S^{Δ} and S^D stimuli responded to and the ratio of these percentages during successive trials. After an initial increase in S^{Δ} responding, the ratio of responses in S^{Δ} to those in S^D falls rapidly. Following the change in instructions after trial 4, there is an increase in the number of S^{Δ}'s responded to, reaching 100 per cent by trial 7. However, the ratio of the number of responses in S^{Δ} to those in S^{D} falls abruptly after trial 4, since the subject only responded once to each S^{Δ} but several times to each S^{D}. The discriminative stimulus that abruptly came to control his vocal behavior was, of course, the reinforcing event. The latency functions for S_3 are similar to those for S_1, and are not presented.

The properties of discrimination learning with rise-time as the stimulus variable are, therefore, similar to those when vowel quality was the stimulus variable. What little change may take place in the frequency and latency of responding to S^{D} and S^{Δ} takes place early in the session. Thereafter, differential responding does not change appreciably. The major difference between rise-time and vowel quality as stimulus variables seems to be only in the higher frequency of S^{Δ} responding to the rise-time stimuli. The effect of the change in the contingencies of reinforcement and instructions during the experimental session was an abrupt increase in both S^{D} and S^{Δ} responding. The partial control over responding exerted formerly by the rise-time variable was now relinquished to the reinforcing event. Following each stimulus, the subject "primed" the schedule with a single response. If he was reinforced, a rapid rate of responding followed; if not, there was a pause in responding until the next stimulus. The development of this discrimination was as rapid and abrupt as that which characterized the initial rise-time discrimination.

3. Discrimination of Formant Onset Time in an Aphasic Subject

Since the temporal properties of formants play a large role in speech recognition, this variable does not normally commend itself for an investigation of initial discrimination learning in human adults. However, we recently had occasion to test an aphasic subject who was observed not to discriminate among the phonemes /d/ and /t/; more extensive measures of his consonant discriminations revealed that we could indeed study the "initial" acquisition of a speech discrimination with this subject. Since the details of this research are reported elsewhere (Lane, 196_), a summary of the procedure and findings will suffice; the present discussion focuses on the course of discrimination learning. A series of seven speech stimuli were prepared at the Haskins Laboratories using the Pattern Playback to convert handpainted spectrograms into sound. The stimuli were identical except for the relative onset time of their first and second formants: the first formant was "cut back" in ten msec. steps from 0 to 60 msecs. When normal adults are instructed to identify the stimuli in this series, presented in random order, they always call stimulus 0 /do/ and stimulus 60 /to/. Our aphasic subject reliably called stimulus 0 /do/ but he also called stimulus 60 /do/ about 85 per cent of the time.

To train the discrimination between /do/ and /to/, S was seated at a table in front of a loudspeaker, a reinforcement light, and two buttons, one labeled "do," and the other "to." He was told that when he pushed the correct button the light would flash. The stimulus sequences shown in Table I were recorded on magnetic tape at three-second intervals and later played back to the subject. The corresponding response sequences are shown at the right of Table I. It will be seen that,whereas stimulus 60 formerly evoked the /to/ response only rarely,it now did so reliably. Abruptly, a "poor" discrimination (relatively little differential responding) showed a marked improvement, although there was some perseveration on the /do/ response during the first few trials.

4. Discrimination of Pure Tone Intensity

As in experiment 3, the data to be reported here were obtained incidentally in the course of an investigation of a different problem: the discriminative control of concurrent responses. The procedure is reported elsewhere (Cross and Lane, 196_) and will only be summarized here. The subject wore a binaural headset while seated in front of a microphone and counter within an anechoic chamber. He was told that by saying /ka/ or /ti/ at appropriate times he could produce points on the counter and that his pay was related to the number of points obtained. One hundred and forty 500 cps tones were presented to the subject in random order, half at 56 db and half at 74 db (SPL). The 56 db tone was the discriminative stimulus ($S^D{}_1$) for a /ti/ response (R_1) and the 74 db tone was the discriminative stimulus ($S^D{}_2$) for a /ka/ response (R_2). $S^D{}_1$ was S^Δ for R_2 and $S^D{}_2$ was S^Δ for R_1. If a single /ti/ response followed $S^D{}_1$ or a single /ka/ response followed $S^D{}_2$ within the allowed time interval (5.5 secs.), reinforcement was provided on each of the first ten occasions. After that a partial reinforcement schedule was employed with probability of reinforcement equal to .30.

Table II shows that all of the 20 subjects emitted their first correct discriminative response within the first three presentations of $S^D{}_1$ and $S^D{}_2$. Subsequently, most of the subjects responded correctly 100 per cent of the time. Thus, 16 of the 20 subjects emitted five correct consecutive responses within eight stimulus presentations. Once again we observe an extremely abrupt change in the discriminative behavior of human adults following a minimum of discrimination training.

5. Discrimination Learning With an Audio-lingual Program

A self-instruction program was prepared to teach discriminations among the vowels and stop consonants of Spanish. The program was subdivided into 14 "frames" each of which comprised approximately 60 stimulus presentations. Each frame included one S^D and several English and Spanish S^Δ's (generally 12 in number), which were presented several times each in irregular order. The 14 frames as well as the 60 stimuli within each frame were "programmed" in the sense that items and frames were sequenced according to a tentative schedule of difficulty for the English-speaking student. The stimuli were rendered by a linguist at approximately four-second intervals and recorded on magnetic tape. In the manner of experiment 1, coding tones were recorded adjacent to each S^D on a second tape track; these tones served during playback to control reinforcement and data recording circuitry. The subject listened to the stimuli with a high-fidelity binaural headset while seated in front of a Lindsley manipulandum and add-subtract counter within an anechoic chamber. He was instructed to respond after the first stimulus in the frame and after all subsequent presentations of that particular sound. If S responded to an S^D or failed to respond to an S^Δ (during the four-second inter-stimulus interval), the counter added one point; if S responded to an S^Δ or failed to respond to an S^D, the counter subtracted one point. The experimenter and all control apparatus were located in an adjacent room where a print-out counter recorded right and wrong responses and reinforcements following each stimulus. Phase one of the experiment constituted a pre-test: each of the three subjects was presented with the tape recorded program but the reinforcement device was disconnected. In phase two the contingencies of reinforcement were in effect. S repeated each frame until he made eight errors or less; he then advanced to the next frame. Phase three constituted a post-test; the procedure was identical to that for phase one.

Table III shows the per cent correct responses before and after the training phase (col. 4). It is clear that within the three hour session we effected an appreciable change in the discriminative repertories of our students. It is equally clear that our tentative program could be revised extensively. The reader will note that the increase in per cent of responses to S^D was greater than twice the decrease in the per cent of responses to S^Δ. This finding is evidence of the subjects' prior discriminative training. It means that our subjects were, initially, over-discriminating and that, during the training phase, they learned to respond correctly to more of the stimulus population.

Most of the phonetic discriminations were mastered in a single trial during phase 2, again showing the transfer of the subjects' prior training. Figure 9 presents the number of frames in which the subjects reached criterion within one through seventeen trials. A third way of describing the relation between conditioning and the change in the discriminative behavior of our subjects is presented in Fig. 10. For each of the 14 frames, the number of errors made during phase two is plotted against the per cent increase in correct responses from pre- to post-test. Evidence of the "hallmark" of discrimination learning—extinction—is lacking. There are some frames in which the number of errors as well as the increase in per cent correct are small; this we infer to be the direct effect of prior discrimination training. There are also several frames in which the number of errors is small but the increase in per cent correct large; this is the discontinuity in auditory discrimination learning to which the present paper is addressed. We infer that this finding also reflects the prior training of our subjects.

A detailed analysis of the data points shown by filled circles in Fig. 10 is given Figs. 11 and 12. Cumulative correct responses to successive stimuli are plotted in Fig. 11 for the case where there were relatively few errors and a large increase in the per cent correct responses from pre- to post-test. It will be seen that S responded to the first S^D (as instructed) but failed to respond to the next two presentations of S^D. Subsequently all responses were reinforced. This limited amount of training, if it may be called that, effected a 47 per cent increase in the number of correct responses from pre- to post-test. Figure 11 is representative of the functions for the other frames in which there were few errors but a marked improvement in discrimination. By way of contrast, Fig. 12 presents data from the same subject for one of the less numerous frames in which there were many errors and an appreciable improvement in discrimination. Here we see the more gradual development of differential stimulus control which is normally associated with the process of discrimination learning.

Summary

When auditory discrimination training is undertaken with human adults the effect of their prior history of reinforcement is certain to play a major role in the course of learning. This history may manifest itself in an initial extent of stimulus control which far exceeds chance levels or it may appear as a discontinuity in the development of differential responding— an abrupt increase to nearly complete stimulus control. There seem to be few, if any, auditory continua that do not sample, at least in part, the subject's prior discriminative repertory. Although the same basic principles apply to discrimination learning in the human adult and in more naive organisms, the present findings suggest that the same conditioning procedures may not be equally efficient.

Footnotes

1. The assistance of Mr. B. A. Schneider and Mr. D. G. Reif are gratefully acknowledged. This research was performed pursuant to a contract with the Language Development Section, U. S. Office of Education.

REFERENCES

Cross, D. V. and Lane, H. L. On the discriminative control of concurrent responses: the relations among response topography, frequency, and latency in stimulus generalization. (submitted to the *J. exp. anal. Behav.*)

Keller, F. S. and Schoenfeld, W. N. *Principles of Psychology*. New York: Appleton-Century-Crofts, 1950.

Lane, H. L. Operant reconditioning of a consonant discrimination in an aphasic. (submitted to the *J. exp. anal. Behav.*)

TABLE I

STIMULUS SEQUENCES AND RESPONSE SEQUENCES DURING CONSONANT DISCRIMINATION TRAINING WITH ONE APHASIC SUBJECT. (D = /do/, T = /to/.)

Stimulus Sequence	Response Sequence	Errors
D D D D D D D D D D	D D D D D D D D D D	
T T T T T T T T T T	D T T T T T T T T T	1
D D D D D D D D D D	D D D D D D D D D D	
T T T T T T T T T T	D T T T T T T T T T	1
D D D D D D D D D D	D D D D D D D D D D	
T T T T T T T T T T	D T T T T T T T T T	1
D D D D D T T T T T	D D D D D D T T T T	
D D D D D T T T T T	D D D D D D T T T T	1
D D D D D T T T T T	D D D D D D T T T T	
D D D D D T T T T T	D D D D D D T T T T	1
D D D D D T T T T T	D D D D D T T T T T	
D D D D D T T T T T	D D D D D T T T T T	0
D D T T D D T T D D	D D T T D D T T D D	
T T D D T T D D T T	T T D D T T D D T T	0
D T T T D T T D D T	D T T T D T T D D T	
T D D T D D T D D T	T D D T D D T D D T	0

TABLE II

DISCRIMINATION OF PURE TONE INTENSITY. THE NUMBER OF SUBJECTS WHO EMITTED THEIR FIRST CORRECT RESPONSE OR FIVE CONSECUTIVE CORRECT RESPONSES AFTER THE NUMBER OF S^D PRESENTATIONS SHOWN. (N = 20)

Number of S^D presentations to first correct response

	1	2	3	4
S^D_1	13	4	3	0
S^D_2	17	1	2	0

Number of S^D presentations to five consecutive correct responses

	5	6	7	8	9	10	11-15	16-20	21-25	26-30	>30
S^D_1	8	3	5	0	1	1	1	0	0	1	0
S^D_2	15	0	0	1	0	0	1	0	1	1	1

TABLE III

PRE-TEST AND POST-TEST RESPONSES TO PHONETIC S^D's AND S^Δ's IN AN AUDIO-LINGUAL PROGRAM

1	2		3		4	
Subject	per cent S^D:R		per cent S^Δ:R		per cent(S^D:R+S^Δ:~R)	
	pre	post	pre	post	pre	post
1	55	76	19	21	67	79
2	61	76	29	19	68	79
3	53	72	34	20	65	82
4	53	77	19	17	70	86

FIGURE CAPTIONS

Fig. 1. Vowel discrimination, S_1. (a) Cumulative per cent of stimuli responded to on each of three trials. (b) Cumulative latency of consecutive responses in S^D and S^Δ during trial 1.

Fig. 2. Vowel discrimination, S_2. (a) Cumulative per cent of stimuli responded to on each of two trials. (b) Cumulative latency of consecutive responses in S^D and S^Δ during trial 1.

Fig. 3. Vowel discrimination, S_3. (a) Cumulative per cent of stimuli responded to on each of seven trials. (b) Cumulative latency of consecutive responses in S^D and S^Δ during trial 1.

Fig. 4. Rise-time discrimination, S_1. Cumulative per cent of stimuli responded to on each of 12 trials.

Fig. 5. Rise-time discrimination, S_1. Cumulative average latency of S^D and S^Δ responses on each of 12 trials.

Fig. 6. Rise-time discrimination, S_1. Cumulative latency of consecutive responses in S^D and S^Δ during trials 1 and 2.

Fig. 7. Rise-time discrimination, S_2. Cumulative per cent of stimuli responded to on each of ten trials.

Fig. 8. Rise-time discrimination, S_3. (a) Cumulative per cent of S^D and S^Δ stimuli responded to on each of seven trials. (b) The ratio of S^Δ to S^D stimuli responded to and the ratio of S^Δ to S^D responses on each of seven trials. After trial 4 the subject was reinforced for every response following each S^D.

Fig. 9. Discrimination learning with an audio-lingual program ($N = 3$). The number of frames in which the subjects reached criterion within one through seventeen trials.

Fig. 10. Discrimination learning with an audio-lingual program ($N = 3$). For each of 14 frames, the number of errors made during discrimination learning is plotted against the per cent increase in correct responses from pre- to post-test.

Fig. 11. Discontinuous discrimination learning with an audio-lingual program. Cumulative correct responses to consecutive stimuli by one subject. S^D is indicated by filled circles, S^Δ by unfilled circles.

Fig. 12. Gradual discrimination learning with an audio-lingual program. Cumulative correct responses to consecutive stimuli by one subject during 12 trials.

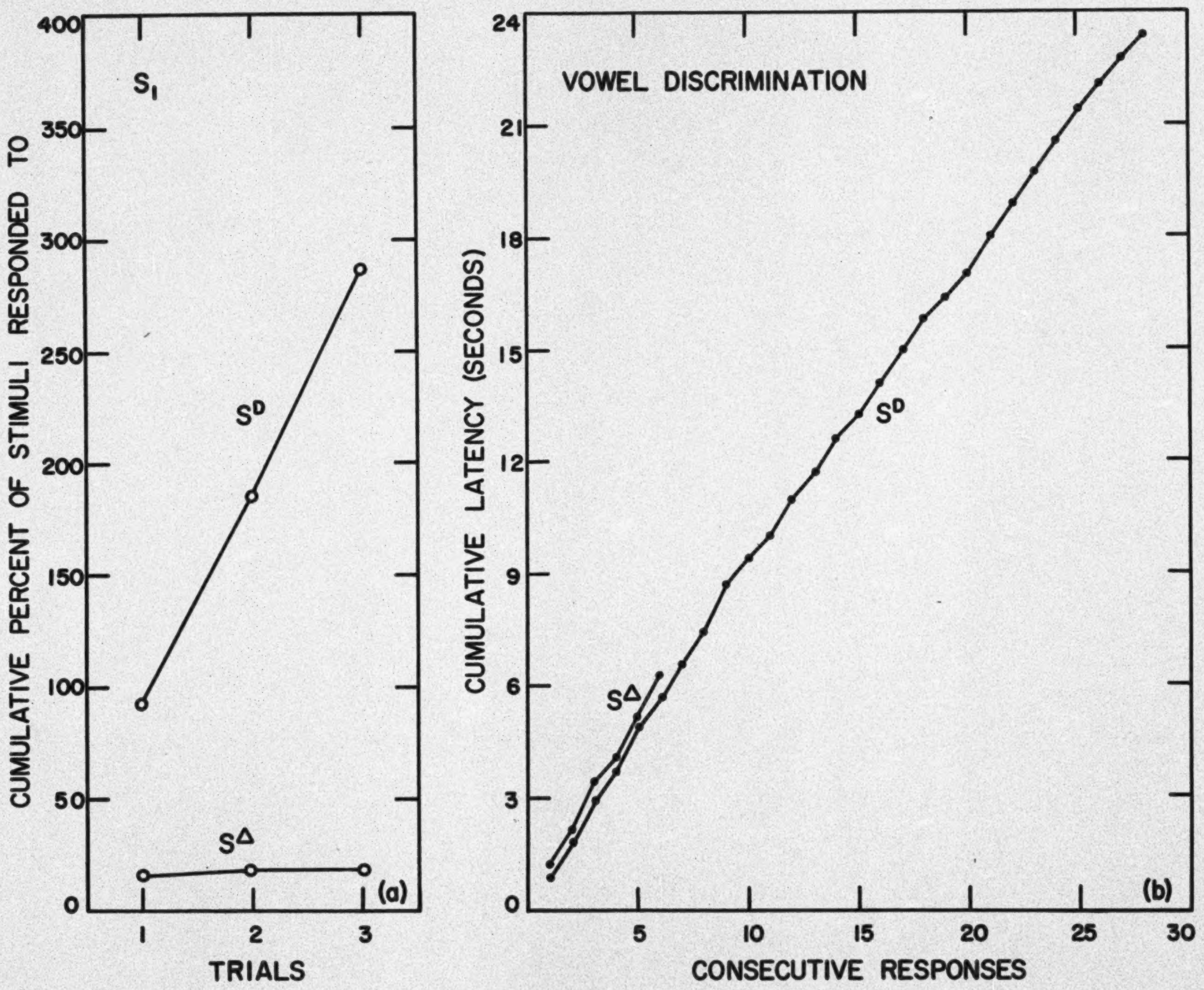
S₁
CUMULATIVE PERCENT OF STIMULI RESPONDED TO
0
50
100
150
200
250
300
350
400
S^D
S^Δ
(a)
1
2
3
TRIALS
VOWEL DISCRIMINATION
CUMULATIVE LATENCY (SECONDS)
0
3
6
9
12
15
18
21
24
S^D
S^Δ
(b)
5
10
15
20
25
30
CONSECUTIVE RESPONSES

Fig. 1

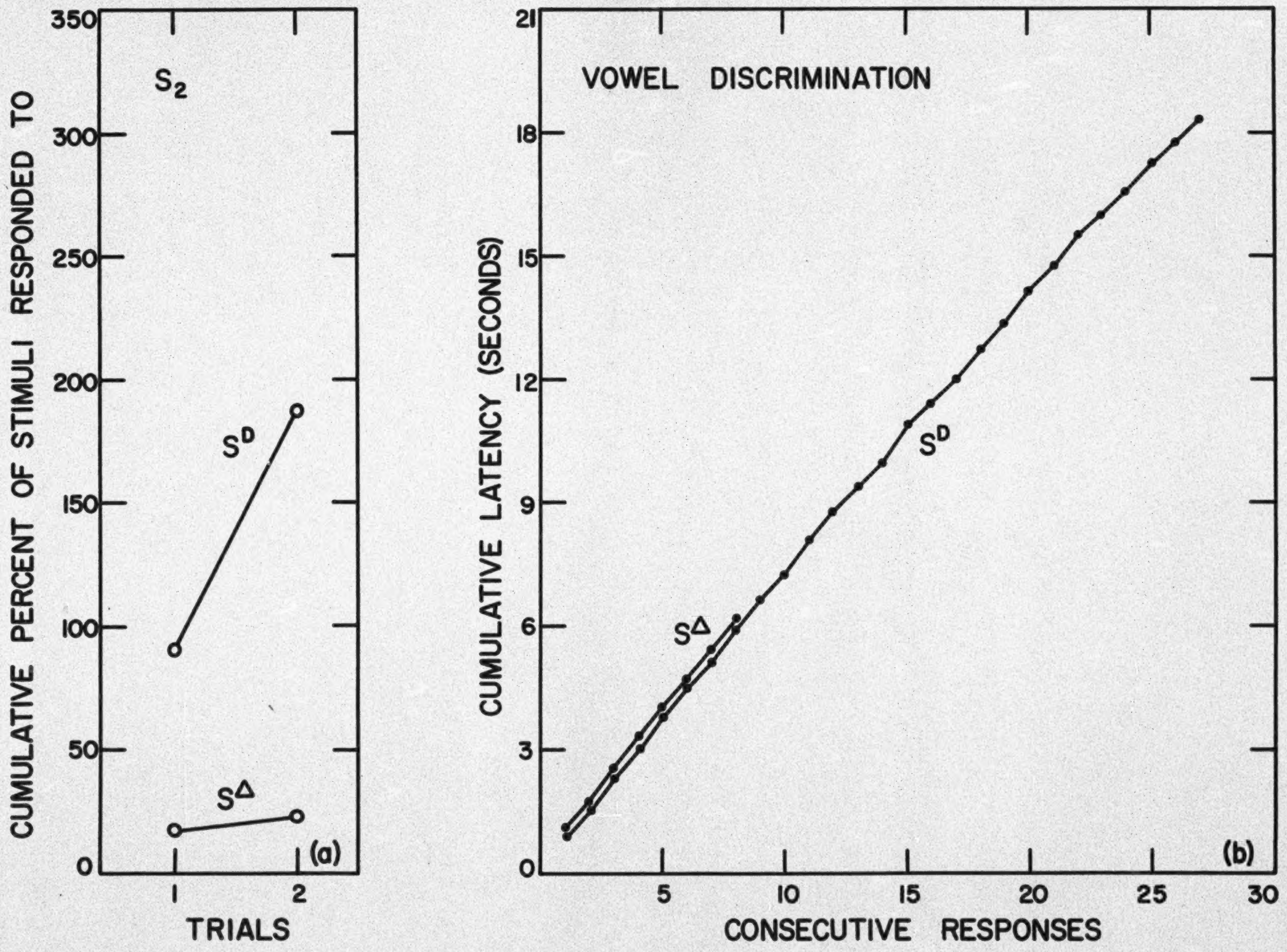
CUMULATIVE PERCENT OF STIMULI RESPONDED TO
350
300
250
200
150
100
50
0
S$_2$
S^D
S$^\Delta$
(a)
1
2
TRIALS
CUMULATIVE LATENCY (SECONDS)
21
18
15
12
9
6
3
0
VOWEL DISCRIMINATION
S^D
S$^\Delta$
(b)
5
10
15
20
25
30
CONSECUTIVE RESPONSES

Fig. 2

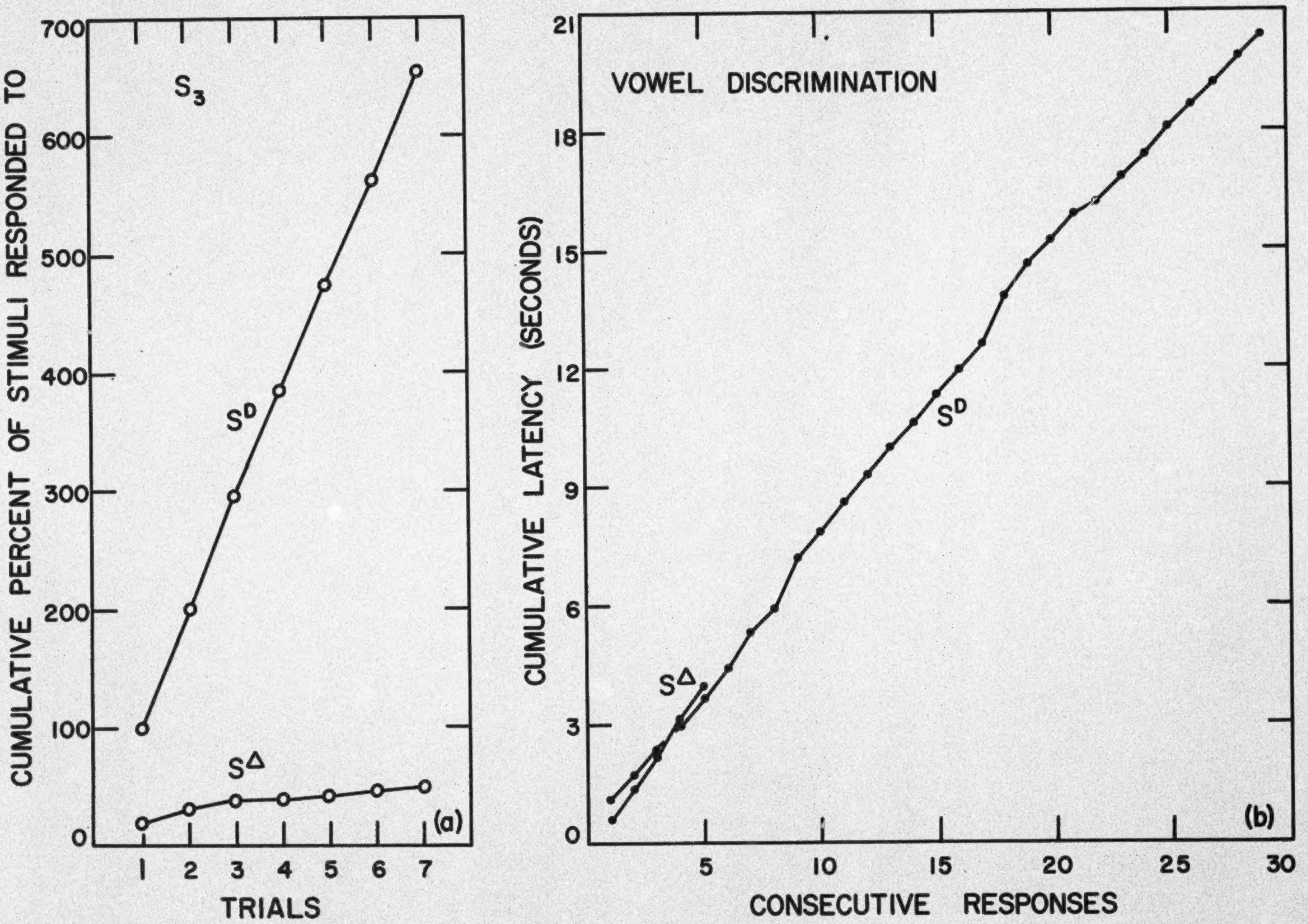
CUMULATIVE PERCENT OF STIMULI RESPONDED TO
0
100
200
300
400
500
600
700
S_3
S^D
S^Δ
(a)
1
2
3
4
5
6
7
TRIALS
VOWEL DISCRIMINATION
CUMULATIVE LATENCY (SECONDS)
0
3
6
9
12
15
18
21
S^D
S^Δ
(b)
5
10
15
20
25
30
CONSECUTIVE RESPONSES

Fig. 3

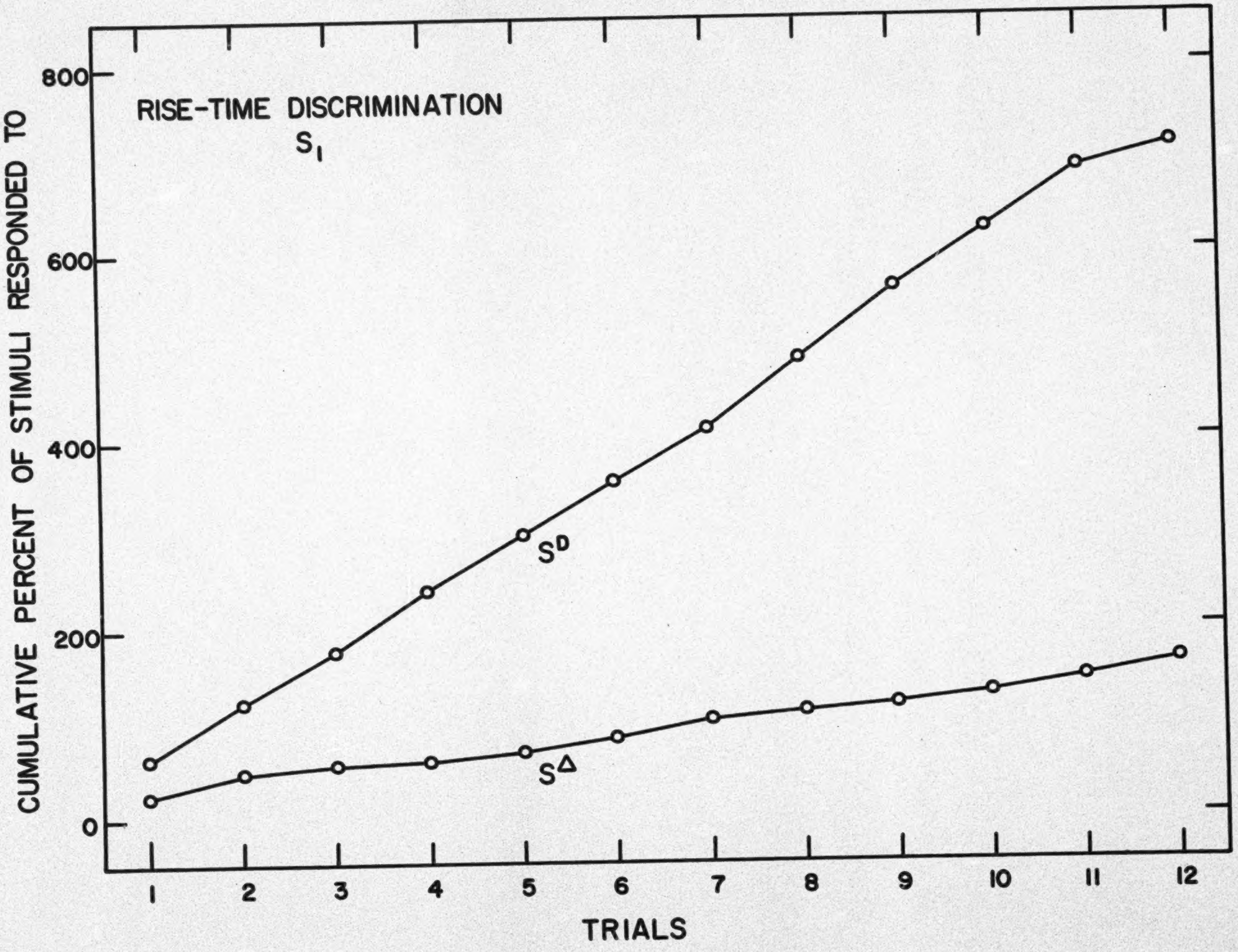
RISE-TIME DISCRIMINATION
S_1
CUMULATIVE PERCENT OF STIMULI RESPONDED TO
800
600
400
200
0
S^D
S^Δ
1
2
3
4
5
6
7
8
9
10
11
12
TRIALS

Fig. 4

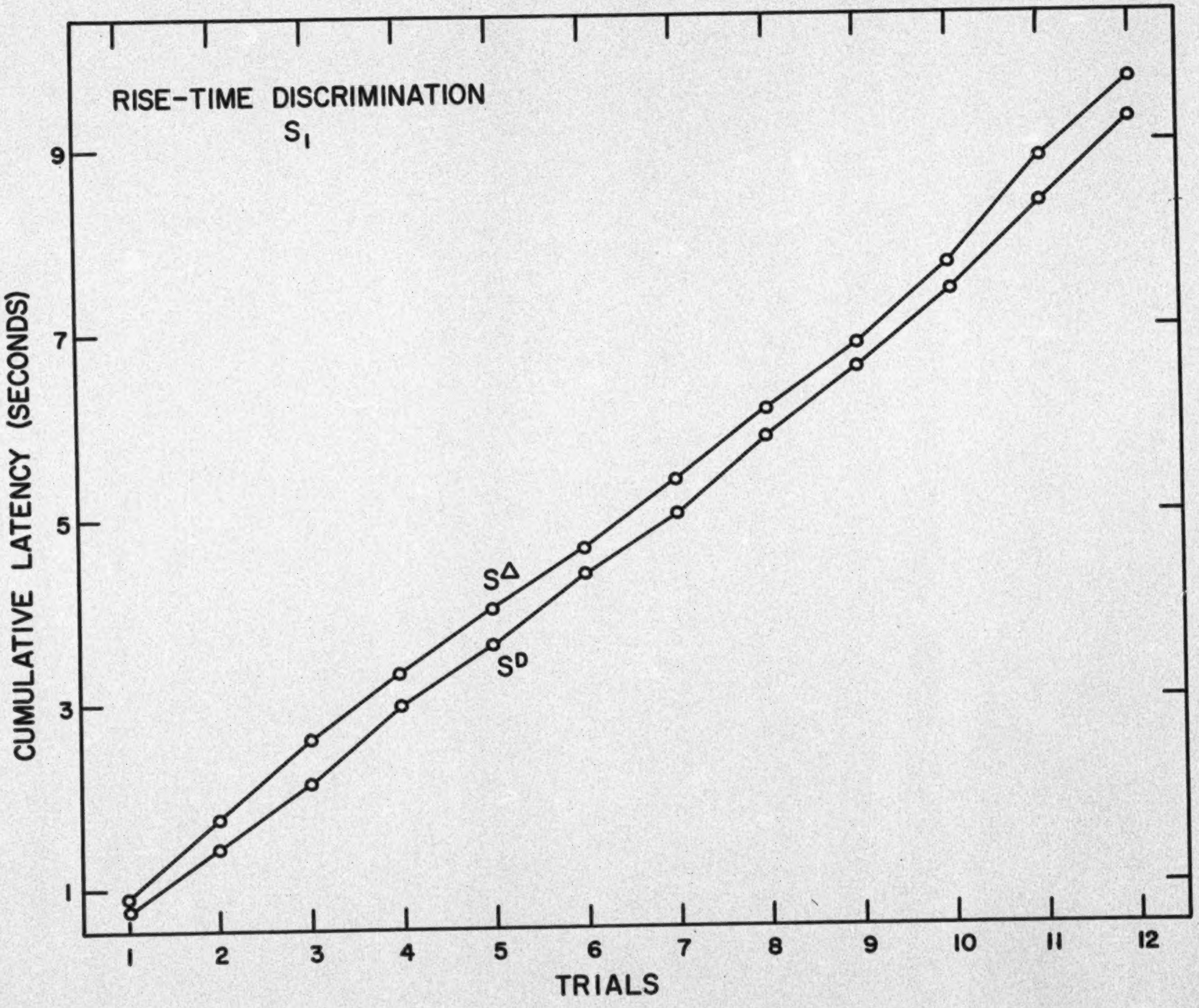
RISE-TIME DISCRIMINATION
S_1
CUMULATIVE LATENCY (SECONDS)
1
3
5
7
9
S^{Δ}
S^{D}
TRIALS
1
2
3
4
5
6
7
8
9
10
11
12

Fig. 5

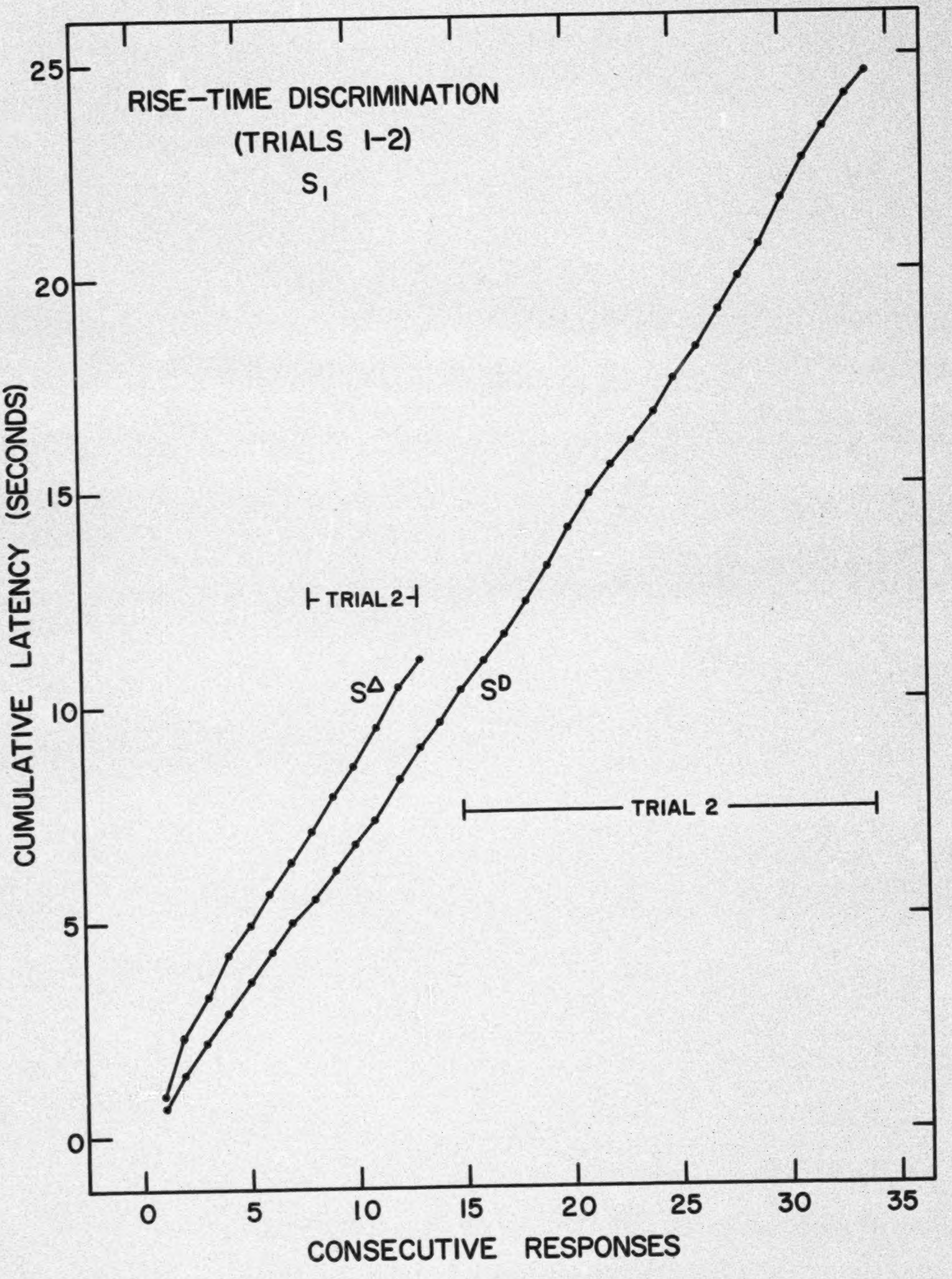
RISE-TIME DISCRIMINATION
(TRIALS 1-2)
S_1
CUMULATIVE LATENCY (SECONDS)
CONSECUTIVE RESPONSES
TRIAL 2
S^Δ
S^D
TRIAL 2
0
5
10
15
20
25
0
5
10
15
20
25
30
35

Fig. 6

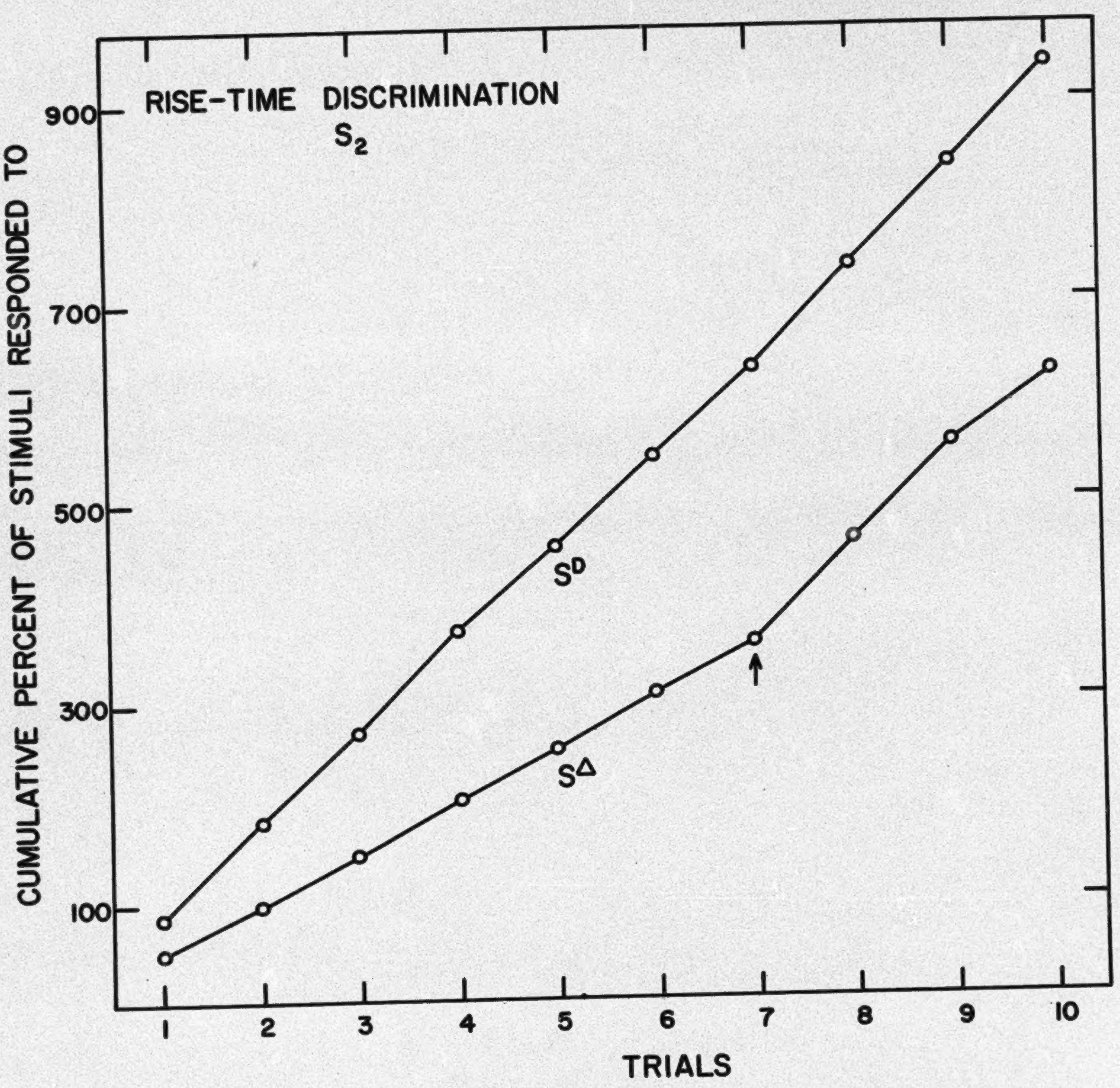
RISE-TIME DISCRIMINATION
S_2
CUMULATIVE PERCENT OF STIMULI RESPONDED TO
900
700
500
300
100
S^D
S^Δ
1
2
3
4
5
6
7
8
9
10
TRIALS

Fig. 7

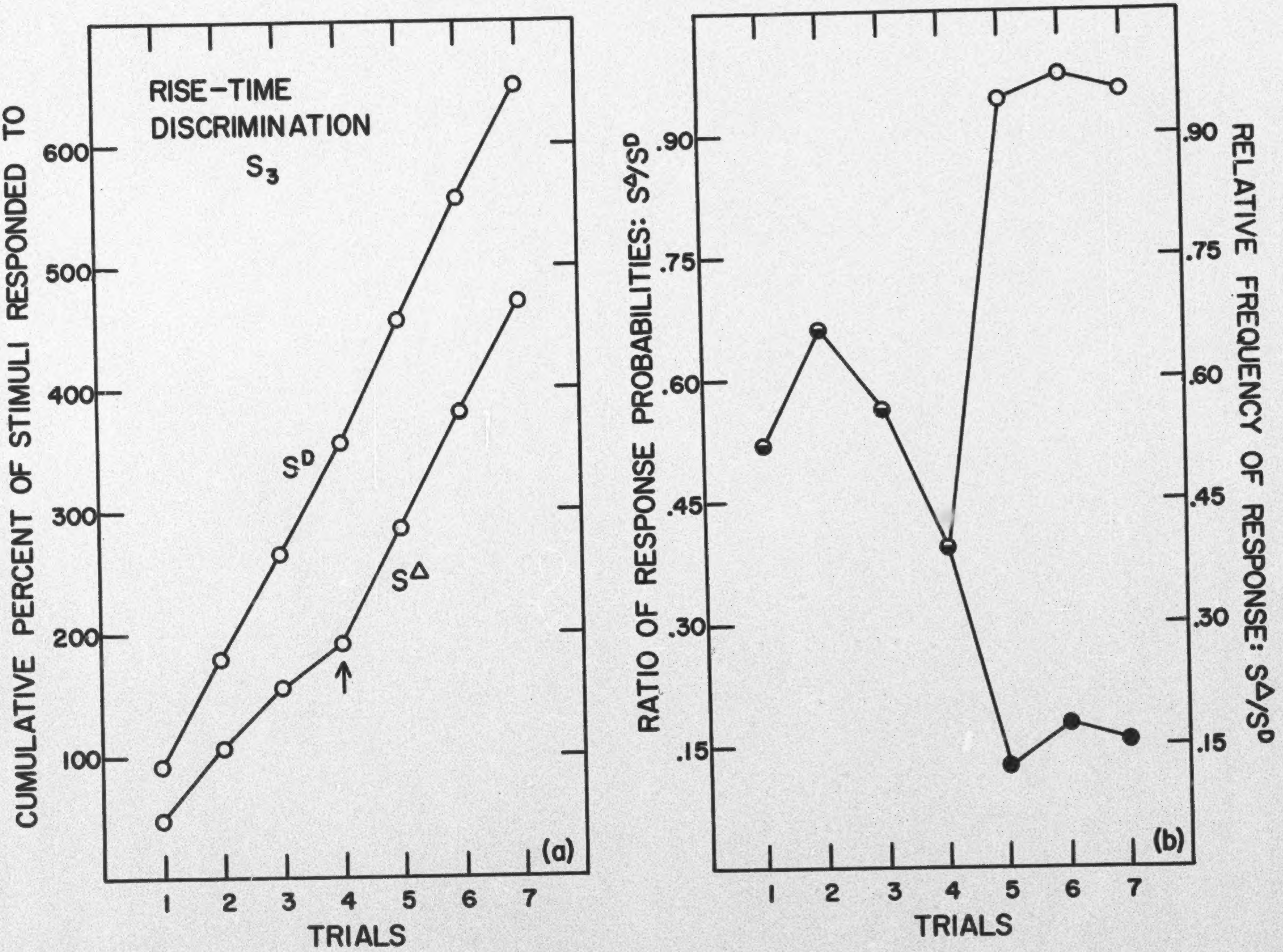
RISE-TIME
DISCRIMINATION
S_3
CUMULATIVE PERCENT OF STIMULI RESPONDED TO
100
200
300
400
500
600
S^D
S^Δ
TRIALS
1
2
3
4
5
6
7
(a)
RATIO OF RESPONSE PROBABILITIES: S^Δ/S^D
.15
.30
.45
.60
.75
.90
RELATIVE FREQUENCY OF RESPONSE: S^Δ/S^D
(b)

Fig. 8

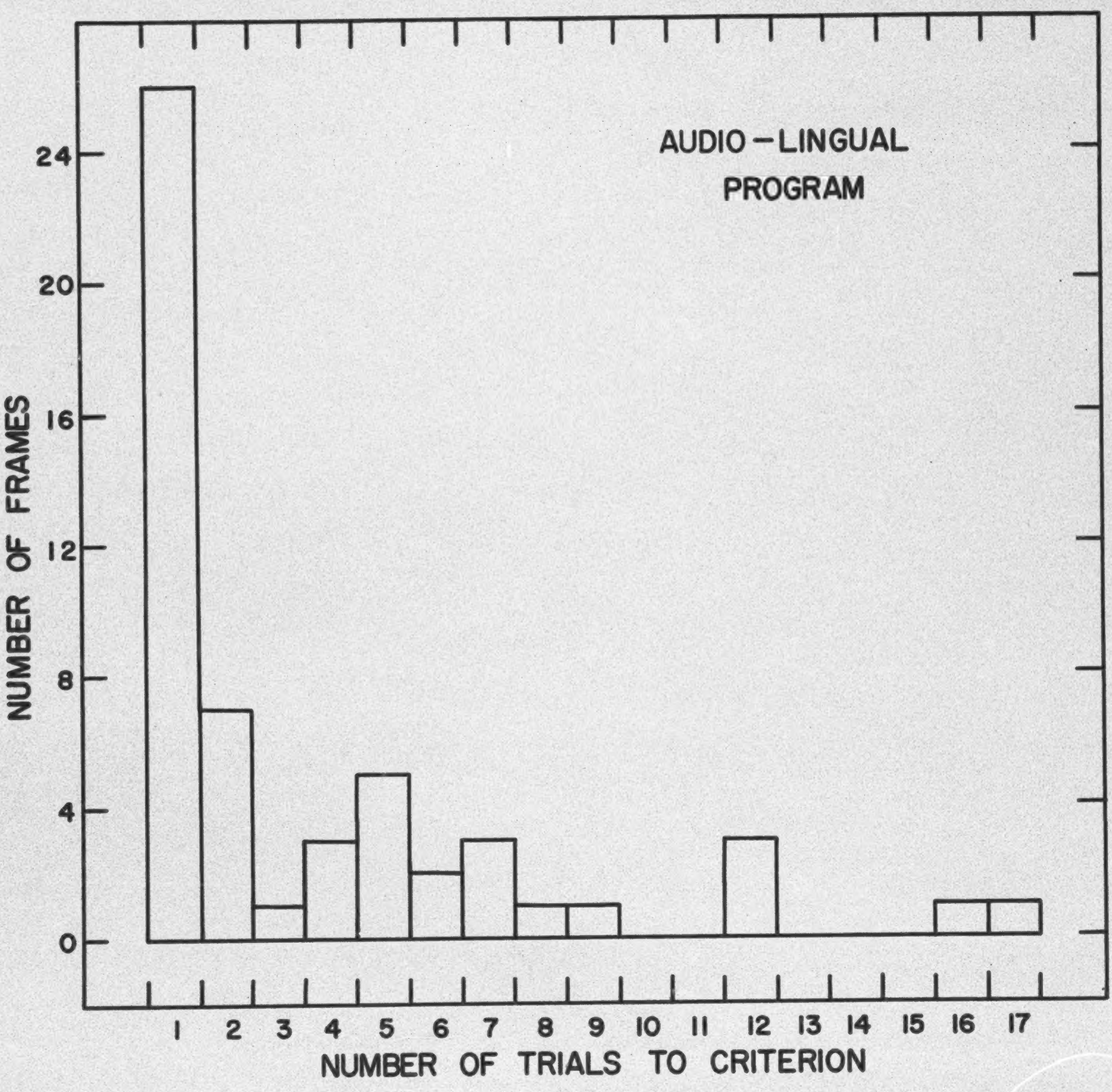
AUDIO-LINGUAL
PROGRAM
NUMBER OF FRAMES
0
4
8
12
16
20
24
NUMBER OF TRIALS TO CRITERION
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Fig. 9

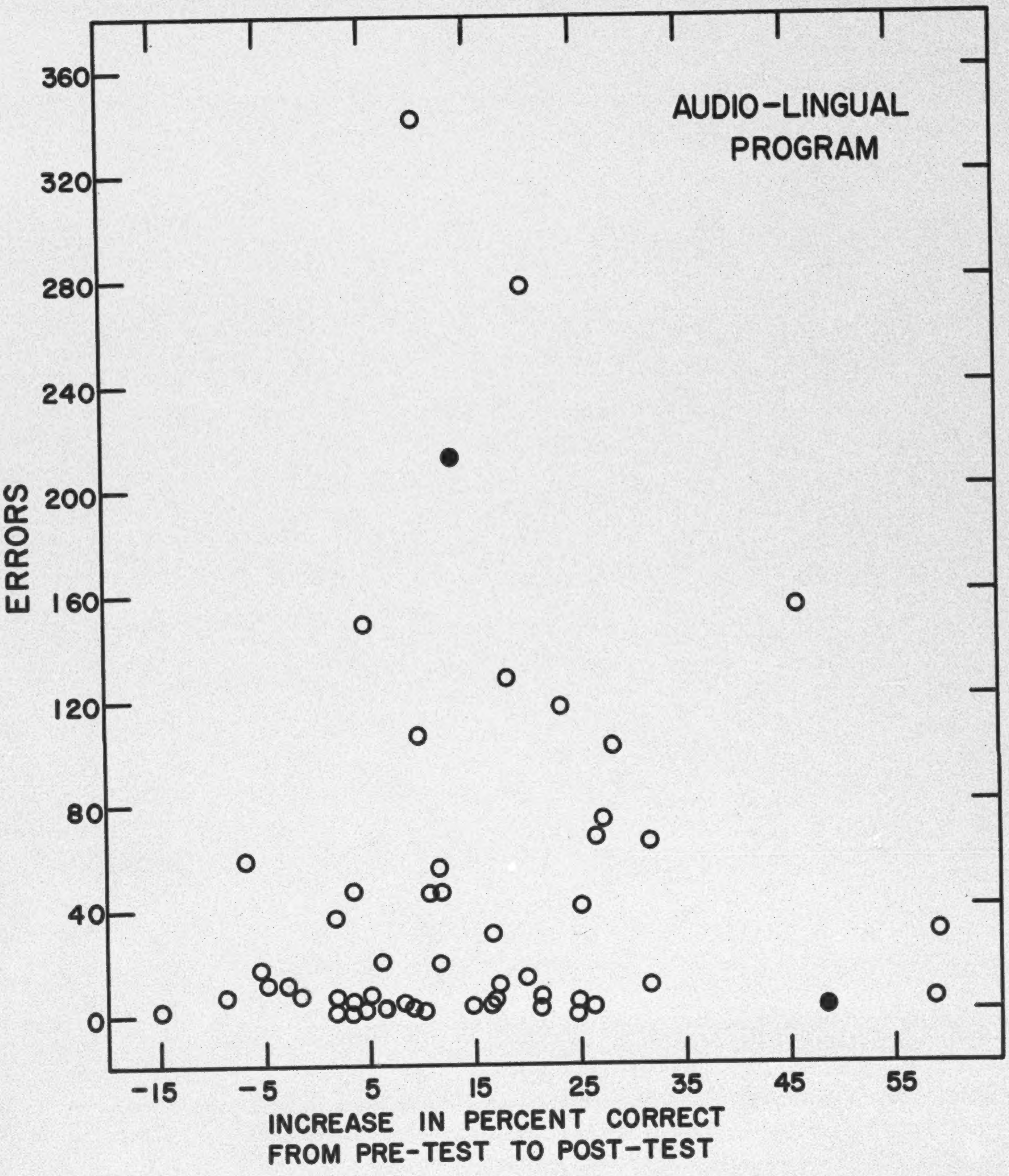
AUDIO-LINGUAL PROGRAM
ERRORS
360
320
280
240
200
160
120
80
40
0
-15
-5
5
15
25
35
45
55
INCREASE IN PERCENT CORRECT FROM PRE-TEST TO POST-TEST

Fig. 10

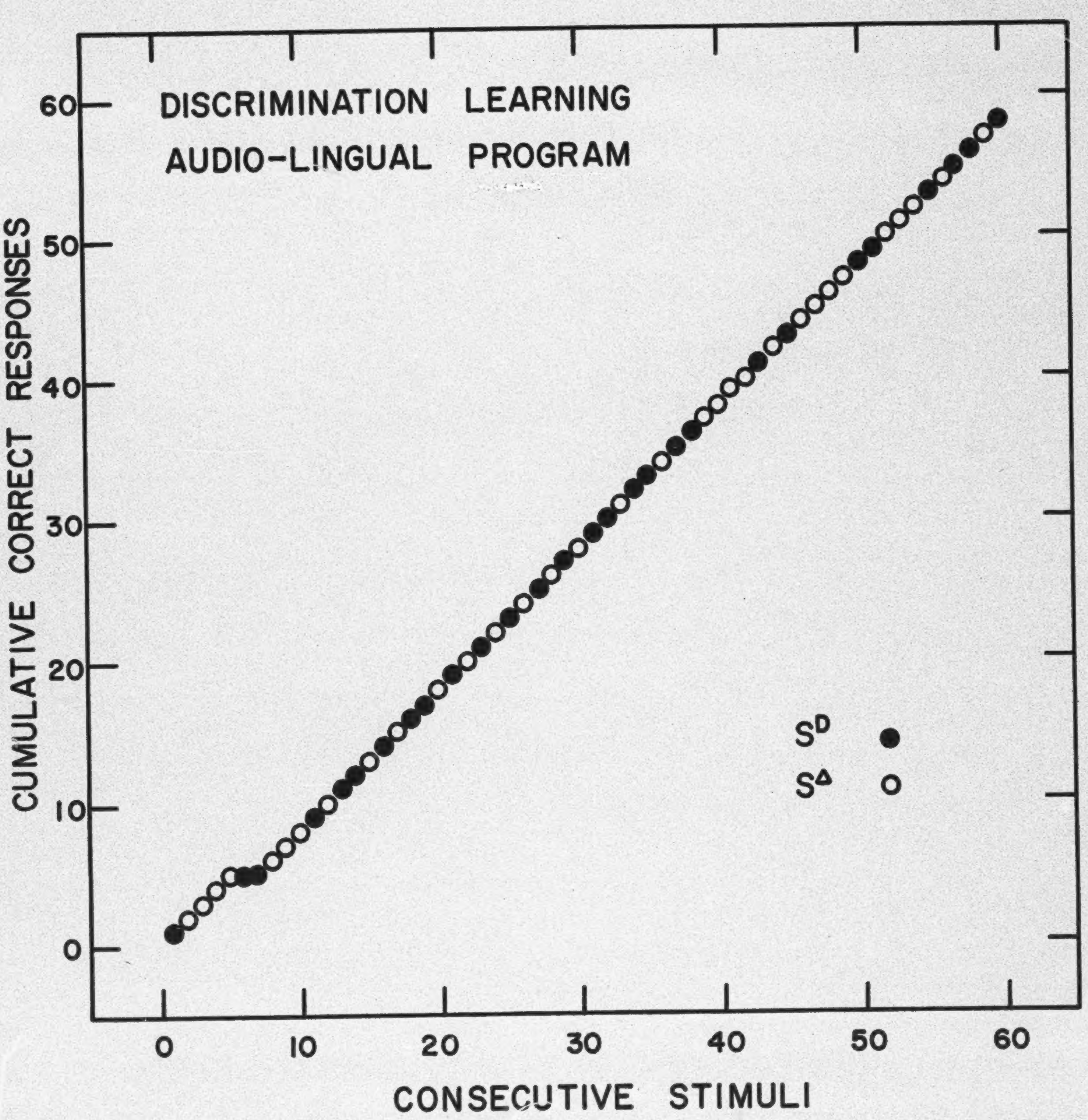
DISCRIMINATION LEARNING
AUDIO-LINGUAL PROGRAM
CUMULATIVE CORRECT RESPONSES
CONSECUTIVE STIMULI
0
10
20
30
40
50
60
S^D ●
S^Δ ○

Fig. 11

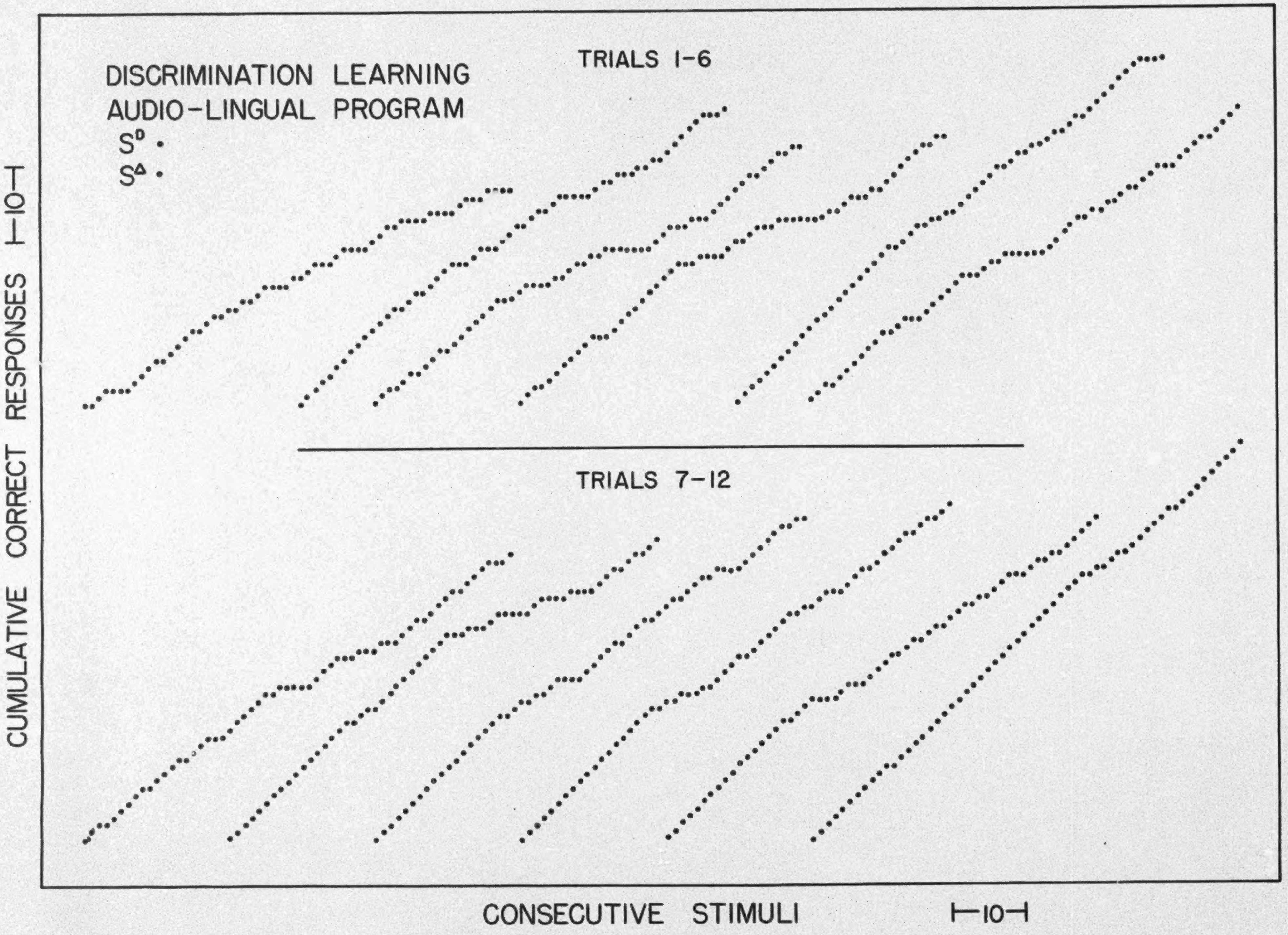
DISCRIMINATION LEARNING
AUDIO-LINGUAL PROGRAM
S^D •
S^Δ •
TRIALS 1-6
TRIALS 7-12
CUMULATIVE CORRECT RESPONSES ⊢10⊣
CONSECUTIVE STIMULI ⊢10⊣

Fig. 12

ON THE DISCRIMINATIVE CONTROL OF CONCURRENT RESPONSES: THE RELATIONS AMONG RESPONSE FREQUENCY, LATENCY, AND TOPOGRAPHY IN AUDITORY GENERALIZATION

D. V. Cross and H. L. Lane
The University of Michigan

ON THE DISCRIMINATIVE CONTROL OF CONCURRENT RESPONSES: THE RELATIONS AMONG RESPONSE FREQUENCY, LATENCY, AND TOPOGRAPHY IN AUDITORY GENERALIZATION

D.V. Cross and H.L. Lane
Communication Sciences Laboratory
The University of Michigan

Studies of stimulus generalization are properly concerned with the changes in behavior effected by changes in the controlling stimulus. The experimental techniques employed in the past to establish stimulus control of responding have been either (1) extensive training in the presence of a single stimulus or (2) discrimination training. In the latter case, responses are reinforced in the presence of S^D, and S^Δ is either the absence of the stimulus or some other value of it on a continuum specified by the experimenter. The operations employed to demonstrate subsequent stimulus generalization have varied among experiments, but, usually, orderly gradients of generalization have been obtained by measuring changes in the rate, frequency of emission, or latency of the response during periods of extinction in which other stimulus values are presented to the organism (Mednick and Freedman, 1960). The gradients obtained show that the degree of discriminative control acquired by a stimulus, S_1, which is not present during training, is a monotonically decreasing function of the distance between S_1 and S^D, measured on a physical continuum.

Prior research, in which stimulus generalization has been studied as a dependent variable, has been solely directed at stimulus generalization following single-response training. It is apparent, however, that discriminative behavior is often acquired by the concurrent conditioning of several responses, each under the control of a different discriminative stimulus. The simplest experimental paradigm appropriate to the investigation of this problem consists of discrimination training with two mutually incompatible responses, each reinforced in the presence of a different S^D. Reinforcement and extinction are reciprocal operations in this procedure in that the S^D for one response also serves as an S^Δ for the other. Two questions that arise are (1) what are the properties of stimulus generalization following this conditioning procedure, and (2) how does this behavior compare with that following single-response conditioning?

The present study[1] answers these questions by examining the changes in probability, latency and topography of human vocal responses brought about by changes in an auditory discriminative stimulus.

Experiment I

In this first experiment, the vocal responses employed were the phonemic clusters [ka] and [ti]. These responses may be termed topographically discrete because the articulatory gestures necessary to produce them involve different parts of the vocal apparatus and the ranges of topographical variation associated with the two responses do not overlap. Topographically distinct responses were selected so that response generalization would be minimal and thus the findings in stimulus generalization would not be confounded. In Experiment II the effect of both types of generalization operating in concert will be examined.

Method

Fourteen male and six female volunteer undergraduates served individually in sessions lasting 40 minutes. The subject was seated in an anechoic chamber in front of a counter, signal light, and microphone. Auditory stimuli were presented monaurally through a binaural headset with calibrated earphones (PDR-8). The stimuli were 500 cps tones, 1.2 secs. in duration, recorded on magnetic tape at three db intervals over a 30 db range. In order to eliminate print-through signals and reduce noise during playback of the tape recording, an electronic switch (Grason-Stadler Model No. 829S119) and a narrow band-pass filter (Dytronics) were interposed between the tape recorder output (Ampex 300-4) and the headphone.

Pulses synchronized with stimulus onset were recorded on a second tape track; these closed the electronic switch, allowing the stimulus to reach the headphone, and also triggered an electronic counter (Hewlett-Packard 522B). The subject's response to the stimulus operated a voice relay (Miratel) which, in turn, stopped the counter. The start-stop interval was read in milliseconds from the counter and taken as the latency. If S failed to respond, the time intervals were automatically terminated after 5.5 seconds by stop pulses recorded on a third track of the tape. All control apparatus was located outside of the experimental chamber.

Procedure. After the subject was seated in the anechoic chamber, the following instructions were read:

"You can earn money by simply saying /ka/ or /ti/ at appropriate times. We can't tell you now when or how these responses should be used. That is for you to learn. All you have to do is wear this headphone and watch the display unit in front of you. You will hear various sounds. Each time you respond appropriately the green light will flash and five points will be added to your score on the counter. You will want to get as high a score as possible because the amount we pay you at the completion of the experiment will be determined by your final score." (Questions were answered by a repeat of the instructions only.)

1. Training. One hundred and forty 500 cps tones were presented to the subject in random order, half at 56 db and half at 74 db (SPL). The 56 db tone was the discriminative stimulus ($S^D{}_1$) for a /ti/ response (R_1) and the 74 db tone was the discriminative stimulus ($S^D{}_2$) for a /ka/ response (R_2). $S^D{}_1$ was S^Δ for R_2 and $S^D{}_2$ was S^Δ for R_1. If a single /ti/ response followed $S^D{}_1$ or a single /ka/ response followed $S^D{}_2$ within the allowed time interval (5.5 secs.), reinforcement was provided on each of the first ten occasions. After that a partial reinforcement schedule was employed with probability of reinforcement equal to .30. The schedule was adjusted, however, to insure that both responses would be reinforced an equal number of times. At the completion of the training phase, the experimenter reentered the chamber.

2. Testing: The subject was told that the experiment would continue as before but with one change. Although the points earned for appropriate responses would "continue to accumulate on the counter in the other room," his own display unit would be inoperative. The counter and signal light were

disconnected and the display moved out of view.

One hundred and ten stimuli were presented in random order at eleven intensity levels (See Fig. 1).

Results

Figure 2 summarizes response probability and latency data for all 20 subjects. Each circle is an estimate of the conditional probability, when stimulus i is presented, of the R_1 response previously conditioned to S^D_1; similarly, the squares give $p(R_2/S_i)$. These estimates are based on the relative frequency of both responses in a sample comprising 200 presentations of each S_i. The R_1 and R_2 probability functions are not exact complements of one another, since S was not forced to respond to each stimulus. The total number of responses emitted to each stimulus by the 20 Ss varied from 190 to 199; the lowest totals occurred at the middle stimulus values. Figure 2 shows that the maximum number of R_1 and R_2 responses occurred at the extreme low and high intensities, respectively, and not at the two S^D intensities (see discussion). The reader will also note an asymmetry of the R_1 and R_2 gradients, showing greater generalization of R_1 to high stimulus intensities than generalization of R_2 to low stimulus intensities.

The average latencies for the two responses combined are shown by the dotted curve. Examination of response latencies reveals minima when the probability of one response was high and the other low. Any change in response probabilities toward equality was correlated with increased latencies; the latency function reaches a maximum when the probabilities of the two responses are most nearly equal. Figure 3 presents a breakdown of the total latencies into those associated with R_1 and R_2. Latencies accompanying the stochastically dominant response (the response with the highest probability of occurence at a given stimulus intensity) are consistently shorter than the latencies associated with the non-dominant response. Both latency functions increase to a maximum at a point displaced 12 db from their respective S^D intensities and then decrease systematically (see discussion).

Since all subjects were given the same amount of discrimination training during the first phase of the experiment and an arbitrary learning criterion was not imposed, it was possible to partition the generalization data with respect to how well the initial discriminations were formed. The subjects were divided into two groups of ten each on the basis of the number of incorrect responses emitted during the second half of the training session, that is, the last 70 stimulus presentations. The number of "errors" (S^D_1: R_2 and S^D_2: R_1) made by the subjects in Group I varied from one to seven with an average of 3.6. The number of errors made by subjects in Group II varied from 10 to 29 with an average of 15.7. A comparison of the R_1 and R_2 generalization gradients for the two groups is presented in Fig. 4. The shapes of the obtained functions are similar; the major differences contrasting the two groups are the greater degree of generalization and the greater number of responses emitted by Group II. (Group I emitted 1051 responses out of a possible 1100 and Group II emitted 1091.) Figure 5 gives the mean response latency to each stimulus intensity for the two groups and reveals a third difference. Group I had appreciably higher maximum and lower minimum latencies than Group II. Comparison of Figs. 4 and 5 shows that the inverse relation between the ratio of response probabilities and the latency at each stimulus intensity holds for each of the two groups as well as for their combined data (Fig. 2).

Experiment II

The preceding experiment employed two vocal responses that were mutually incompatible and topographically discrete. In the present experiment the basic conditions of Experiment I were replicated. However, in order to examine the possible effects of stimulus-response interaction, two vocal responses were employed that were topographically continuous. These responses differed only with respect to fundamental frequency, the acoustic correlate of a topographical continuum (tension on the vocal cords) along which response generalization may be observed and conveniently measured.

Method

Fourteen male students, none of whom participated in the preceding experiment, were subjects. Apparatus and procedure were basically the same as in Experiment I with the following exceptions. A pitch meter and graphic level recorder (General Radio Co. Type 1521- A) were used to measure and control the fundamental frequency of the responses emitted by S. The former device consisted primarily of a series of filters, frequency scanning circuits, and electronic switches which permitted selection of the fundamental frequency of the vocal response from the complex speech signal, and a frequency meter (Hewlett-Packard Mod. 500 BR) which transformed a sinusoidal input into a d-c output voltage proportional to the input frequency. The d-c output of the meter was applied to the graphic level recorder for an instantaneous, real-time display of the pitch level of the emitted response.

Procedure. The procedure differed from that of the preceding experiment in that a lengthy session for shaping the desired responses was necessary before discrimination training could begin. The subject was seated in the anechoic chamber and given the following instructions:

> "This is an experiment in pitch production. We want you to learn to produce two levels of vocal pitch by humming. You will learn these pitches by producing a steady and continuous hum and maintaining it until one of the lights in front of you flashes on. If the middle green light flashes, that will indicate that you have produced a correct pitch. You should stop and repeat it. If the top red light flashes your pitch is too high. You should stop and produce another pitch at a lower level. If the bottom yellow light flashes your pitch is too low. You should stop and try a higher pitch. We will start with one pitch level and work with it until you can produce it repeatedly without error, then we will switch to the other pitch. When you have learned to produce that one correctly we will alternate systematically from one to the other to give you practice on both. How well you learn to produce these pitches will help you later in the experiment to win money."

The two vocal pitches required of each S were 147 cps and 227 cps. A pitch production within ± 2 cps of that desired was reinforced.

If after one hour of shaping, the subject was unable to produce reliably the pitches desired, he was excused from the experiment. If the pitches were produced to a criterion of ten successful alternations, shaping procedures were terminated and discrimination training was begun. From this point on the procedure followed that of the preceding experiment. The instructions given the subject were the same except pitch level was substituted for /ka/ or /ti/ response. The lights signaling that the produced pitches were too high or too low were not used. Only the green light and the addition of 5 points to the subject's score signaled a correct response.

The discriminative stimuli for the vocal responses were recorded at the same sound pressure levels as those in Experiment I. Instead of 500 cps tones, however, the stimuli were narrow band noises, with center frequency 5,000 cps, duration 1.2 seconds. Noise rather than tone was employed because the 500 cps tones tended to produce changes in vocal pitch toward matching at 125 cps or 250 cps. Testing for generalization was carried out along the same intensity range as before but with noise instead of tone stimuli.

Results

Of the twelve subjects who started in the experiment, seven satisfied the shaping criterion and continued into the discrimination training phase. Of these seven, four subjects failed to emit one or the other of the differentiated pitches in the presence of the discriminative stimuli and were excused from the experiment. For the remaining three Ss, the responses emitted during testing were analyzed with respect to variations in pitch. These were distributed between two response categories referred to here as low, (R_1), and high, (R_2), pitch productions. It is possible to categorize the pitch continuum in this way because these categories delimit two regions which are separated by an extended range within which no pitches were produced.

The results are presented separately for the three Ss in Fig. 6. The median frequency (circles) and the range (vertical lines) of high and low pitch responses are represented as a function of stimulus intensity. The dotted horizontal lines in each graph represent the absolute pitch levels differentiated in the preceeding training session. Although these absolute levels were not accurately maintained by two of the Ss, the ratio of R_1 to R_2 pitch remains the same as that during training.

The probability of an R_1 response at each stimulus intensity is also shown in Fig. 6 for each of the three subjects. There were no response omissions in this experiment: therefore, the R_2 function is the exact complement of the R_1 function for each S and is not shown. The functions labelled L_T represent the average latency of the responses emitted at each stimulus intensity.

In general the results are in accordance with those of experiment I. The latencies vary systematically with the probability of response functions--- tending toward a maximum where response probabilities are nearly equal and a minimum where response probability is unity or zero. As reported earlier, response dominance occurred at the extremes of the stimulus continuum and the generalization from low to high intensity of the stimulus continuum was greater than that from high to low.

Experiment III

The preceding experiments employed discrimination training procedures in which discriminative responses were reinforced under controlled conditions.

In the present experiment no attempt was made to condition discriminative behavior prior to generalization testing. The two vocal responses /do/ and /to/ were used. It was presumed that these responses were extant in the vocal repertory of the subject and also that, during prior verbal learning, the acoustic patterns correlated with these responses had acquired some discriminative control over the responses themselves. One property that distinguishes the acoustic patterns correlated with /do/ and /to/ is the relative onset time of their first and second formants. This variable defined a stimulus continuum which was sampled at seven points by means of speech synthesis techniques.

Method

In condition (a) of the experiment, S was instructed to respond with /do/ upon hearing the "/do/ stimulus" and /to/ upon hearing the "/to/ stimulus." In order to demonstrate a possible interaction between previously conditioned discriminative responses and competing responses introduced in the experimental situation, a second condition (b) was studied in which the subjects were instructed to reverse their discriminative responses. That is, instead of responding with /do/ to a "/do/ stimulus" they were to respond with /to/, and, accordingly, to respond with /do/ to a "/to/ stimulus". In addition, a third condition (c) was studied in which /ka/ and /ti/ were substituted for the /do/ and /to/ responses. Presumably, this latter procedure would have the effect of introducing multiple competing response tendencies at stimulus values intermediate to the two basic speech sounds.

To obtain generalization gradients of frequency and latency for these responses, seven synthesized speech sounds were prepared using the Pattern Playback[2] to convert hand-painted spectrograms into sound. The spectrographic patterns used, shown in figure 7., were identical except for the relative onset time of their first and second formants: the first formant was "cut back" in 10 millisecond steps from 0 to 60 msecs. Liberman _et_ _al_. (1961) have shown that, with normal adults, the relative frequency of /do/ responses decreases as first formant cutback is increased.

Procedure. The apparatus and procedure were similar to that of Experiment I. Six subjects, undergraduate students who did not participate in the previous experiments, were run individually for sessions lasting 90 minutes. The subject was seated in an anechoic chamber and read the following instructions:

> "When you put on the earphones you will hear a series of sounds which resemble either /do/ or /to/. When you hear /do/ call it __(R_1)__. When you hear /to/ call it __(R_2)__. Always respond to each sound."

In condition (a) the responses requested as R_1 and R_2 were /do/ and /to/ respectively. In condition (b) they were /to/ and /do/, and in condition (c) they were /ka/ and /ti/. Each subject served under all conditions, presented in counterbalanced order so that all permutations of the three conditions occurred.

Results

The results were highly consistent across individual subjects and no systematic effect was observed related to the order in which the three conditions were imposed. Therefore, the data were pooled and the results analyzed on the basis of group totals. Overall, the results replicate the findings of the preceding experiments. As shown in Fig. 8, the gradients of response probability (representing the relative frequency of R_1 at each stimulus value) were not substantially different for the three conditions. Here, as in Experiment II, the R_2 gradients were exact complements of their respective R_1 gradients. The major "between conditions" effect was revealed in the analysis of response latencies. Figure 8 shows the average response latency at each stimulus value for the three conditions employed. It is apparent that the overall latencies in conditions (b) and (c) were substantially longer than those obtained in condition (a). The general shape of the latency functions, however, were similar. Separate analysis of the R_1 and R_2 latencies revealed, as shown in Fig. 9, the same effect observed in Experiment I, i.e., the tendency for response latency to increase systematically then decrease as the test stimulus was changed.

Discussion

The outcome of these experiments suggests that the principles formulated in single-response studies of generalization may be extended on several counts to the multiple-response case. It was observed in Experiment I that the maximum number of R_1 and R_2 responses occurred at the extreme high and low intensities, respectively, and not at the two S^D intensities. This finding is comparable to that reported by Pierrel and Sherman (1960) in a study of the generalization of auditory intensity following discrimination training with a single response. The authors attributed this finding to the generalization of extinction effects resulting from the preceding discrimination training. They suggested that the effects of extinction on S^{Δ} responding may extend along the stimulus continuum as far as, and beyond, S^D. These effects presumably interact with the generalization of effects due to reinforcement in S^D and produce a displacement of response gradients away from S^{Δ}. Hanson (1959) was the first to systematically demonstrate this phenomenon. He showed that the magnitude of the displacement effect is systematically related to the distance (on a physical scale) between S^D and S^{Δ} during training.

In both Experiments I and II there was observed greater generalization of R_1 responses to stimuli more intense than S^{D_1} than there was generalization of R_2 responses to stimuli weaker than S^{D_2}. In addition, the latencies of the R_1 responses to high-intensity stimuli were, overall, shorter than the latencies of the R_2 responses to low-intensity stimuli. These findings are comparable to those reported by several investigators (Brown, 1942 and Heyman, 1957) employing single response procedures; they have been discussed by Hull (1949) in terms of an interaction between "stimulus-intensity-dynamism (V)" and "stimulus-intensity generalization." It is interesting to note that, when the stimuli employed did not vary in intensity, the asymmetry was not observed: in Experiment III the generalization of R_1 was symmetric with that of R_2 in the gradients of both response probability and latency.

On the other hand, there were two major findings in these studies which suggest that the detailed properties of stimulus control in multiple-response situations differ from those in the single-response case. In the first place, the probabilities associated with a specific response were maximal over a large range of stimulus intensities and then decreased rapidly as the alternative response became dominant. This contrasts with the generalization gradients usually obtained following single-response discrimination training, which have been described as depicting the "exquisitely precise tuning of the animal to [a particular] aspect of its environment" (Guttman, 1956). Most investigators have found discontinuous generalization gradients that peak at the training stimulus and decay at an exponential rate on both sides. Consequently, the exponential decay function has become the favored expression for describing the process whereby other stimuli acquire discriminative control over the response (Hull, 1953,Shepard, 1957). A generalization suggested by the present results is that multiple-response discrimination training effectively divides the stimulus continuum into sharply defined, response-specific categories or classes. This formulation receives support from another quarter. In a review of research in the area of speech perception, Liberman (1957) reported that subjects identified speech sounds in such a way as to divide the acoustic continuum into sharply defined categories. This is an alternate way of describing the data of Experiment III in the present study. It should be noted that stimulus control may be "categorical" with respect to nominally scaled response events (e.g.,occurrence versus non-occurrence of a response in a given unit of time) and still yield orderly variations in other measures of responding, such as rate, amplitude, or latency. In Experiment II, for example, response latency increased while response probability was unity or zero over several stimulus values.

A second departure from prior findings is provided by the gradients of response latency observed in the present study. Le Ny (1957) and Schlosberg and Solomon (1943) have reported that the time interval between stimulus onset and response is a monotonically increasing function of the difference between the test stimulus and the training stimulus. Moreover, response latency has generally been presumed to be inversely related to response probability. These generalities hold for only a restricted portion of the stimulus continuum in a two-response situation. In both Experiments I and II, the response latency functions exhibited an unexpected discontinuity, and a change in the sign of their slope, at a stimulus value just beyond the middle stimulus. This "distortion" of the latency function must reflect the influence of a factor other than the generalization of the effects of reinforcement in S^D. A description of the effects of this factor may be obtained if two assumptions are made: (1) the effects of reinforcement generalize in like manner for both responses; (2) by averaging latencies for all responses emitted at a given stimulus value, these effects balance out and the resultant form of the latency function represents the effects of this additional factor. The factor is then found to be maximally effective at the stimulus values where response probabilities were most nearly equal.

Two major findings emerge from an analysis of the relations between response probability and latency: (1) the mean latency for all responses at a given stimulus value varied as an approximately linear function of the variance associated with the obtained distribution of response probabilities and, (2) the latency of the stochastically dominant response was consistently shorter that that associated with the non-dominant response. A psycho physical study reported by Kellogg (1931) corroborates these findings. Kellogg used seven fixed pairs of visual intensities as stimuli; in three of these the left half of the visual field was objectively the darker, in three the right half was the darker, and in the remaining pair the fields were equal. In one condition of the experiment, S responded with either "left side darker" (R_1) or "right side darker" (R_2) to each stimulus pair. Figure 10 presents the probability and mean latency of R_1 and R_2. (The relative intensity values for each stimulus pair were not reported, so the stimulus scale is ordinal and any monotonic, increasing transformation of the data yields a possible representation.) Concerning the relations between response probability, latency, and topography, Fig. 6 shows that there is no systematic change in response topography correlated with the changes in probability and latency discussed earlier. This finding is contrary to an expectation presented by Levine (1960). This author suggested that, if a response can be differentiated along a continuum on the basis of some topographical property (in the present case, fundamental frequency), then responses intermediate to the two conditioned discriminative responses may be emitted when stimuli intermediate to the two S^D's are presented. It may be that topographical continuity is a necessary but not sufficient condition for the kind of "response blending" that Levine describes. In the present experiment, the two responses were indeed sampled from a response continuum but they were mutually incompatible. The use of compatible responses might reveal changes in topography correlated with stimulus generalization.

The generalization gradients obtained when stimulus generalization sampled responses from the prior verbal repertory of the subject did not differ noticeably from those obtained following discrimination training in the experimental situation. However, Fig. 8 shows that, when speech stimuli are employed, the choice of the response pair affects the observed latency distribution. Under instructions to respond with /do/ and /to/ to stimuli that are typically called /do/ and /to/, respectively, the Ss gave shorter response latencies than when they were instructed to call these stimuli /to/ and /do/ or /ka/ and /ti/. One way of interpreting this difference in the latency distributions is to say that approximately 20 years of intermittent discrimination training with speech stimuli have yielded a reduction in response latency of approximately 100 msec.

Summary

Auditory generalization gradients of response probability and latency were obtained from human Ss following discrimination training with two vocal responses conditioned to acoustic stimuli of 56 and 74 db (SPL) under conditions in which: (I) the stimuli were 500 cps tones of 1.2 seconds duration and the responses were the phoneme clusters /ka/ and /ti/ and (II) the stimuli were bursts of noise, 1.2 seconds in duration, and the responses were nasalized phonemes, differentiated with respect to fundamental frequency (147 and 227 cps). In addition, generalization was studied under a third condition (III) in which no discrimination training was administered in the experimental situation: instead the prior verbal training of the subject was sampled by presenting the synthetic speech stimuli /do/ and /to/.

Stimulus generalization was observed in conditions (I) and (II) by presenting 11 stimulus intensities varying in 3 db steps from 50 to 80 db SPL, and in condition (III) by presenting 7 speech stimuli varying with respect to relative onset time of their first and second formants over the range 0 to 60 milliseconds. The results under the three conditions were similar. The response probabilities were maximal over several stimulus values at the extreme ends of the stimulus continuum then dropped sharply at stimuli intermediate to the two S^D's. In conditions (I) and (II) there was greater generalization of the response conditioned to the 56 db stimulus to more intense stimuli than there was generalization of the response conditioned to the 74 db stimulus to less intense stimuli.

An analysis of the latencies of the two responses, taken separately and combined at each stimulus intensity, revealed: (a) an increase in latency as the difference between the test stimulus and the initial S^D increased, and (b) a sharp discontinuity in the latency gradient and reversal in trend at intermediate stimulus intensities. The latencies based on total responses were inversely related to the relative frequency of the two responses at each stimulus value. Where the two responses were most nearly equal in probability, latencies were maximal; when one response had unity or zero probability, latencies were minimal. The latencies associated with the stochastically dominant response at a given stimulus value were consistently shorter than those of the non-dominant response.

The relations among stimulus value, response probability, and response latency remained invariant under changes in two parameters of the experiment. The topographically continuous pair of vocal responses (Experiment II) gave essentially the same generalization gradients as were obtained with topographically discrete phoneme clusters. There were no changes in response topography correlated with the characteristic changes in probability and latency during stimulus generalization. When discrimination training was omitted from the experimental procedure and stimulus generalization was measured along a synthetic speech continuum (Experiment III), the response probability and latency gradients observed were comparable to those of Experiments I and II.

References

Brown, J.S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1957, 53, 239-248.

Guttman, N. The pigeon and the spectrum and other perplexities. *Psychol. Rev.*, 1949, 56, 67-76.

Hanson, H.M. Effects of discrimination training on stimulus generalization. *J. exp. Psychol.*, 1959, 58, 321-334.

Heyman, W. Certain relationships between stimulus intensity and stimulus generalization. *J. exp. Psychol.*, 1957, 53, 239-248.

Hull, C.L. *A behavior system*. New Haven: Yale Univ. Press, 1953.

Hull, C.L. Stimulus intensity dynamism (V) and stimulus generalization. *Psychol. Rev.*, 1949, 56, 67-76.

Kellogg, W.N. The time of judgment in psychometric measures. *Amer. J. Psychol.*, 1931, 43, 65-86.

Levine, G. Stimulus-response generalization with discrete response choices. *J. exp. Psychol.*, 1960, 60, 23-29.

LeNy, J-F. Généralization d'une attitude dans une 'épreuve de temps de réaction. *Année psychol.*, 1957, 57, 11-21.

Liberman, A.M. Some results of research on speech perception. *J. Acoust. Soc. Am.*, 1957, 29, 1, 117-123.

Liberman, A.M., Harris, K.S., Kinney, J., and Lane, H. An effect of learning on perception: the discrimination of relative onset time of the components of certain speech and non-speech patterns. *J. exp. Psychol.*, 1961, 61, 5, 379-389.

Mednick, S. A. and Freedman, J.L. Stimulus generalization. *Psychol. Bull.*, 1960, 57, 169-200.

Pierrel, R. and Sherman, J.G. Generalization of auditory intensity following discrimination training. *J. exp. anal. Behav.*, 1960, 3, 313-322.

Schlosberg, H., and Solomon, R.L. Latency of responses in a choice discrimination. *J. exp. Psychol.*, 1943, 33, 22-39.

Shepard, R.N. Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika,* 1957, 22, 325-245.

Footnotes

1. This research was performed persuant to a contract with the U. S. Office of Education, Language Development Section.

2. For a description of the Pattern Playback and its use, see Cooper, F.S., Spectrum Analysis. *J. Acoust. Soc. Am.*, 1950, 22, 761-762.

Figure Captions

Fig. 1. Stimulus intensities used in experiments I and II.

Fig. 2. Conditional probabilities of R_1 and R_2 and the average of their combined latencies at each stimulus intensity. The conditional probabilities were estimated from the total number of /ka/ responses (squares) and /ti/ responses (circles) emitted in 10 presentations of each stimulus intensity to each of 20 Ss. The total latency (hexagons) at each stimulus intensity is the unweighted mean of the average latency of responding by each of 20 Ss.

Fig. 3. Comparison of the two response latency gradients relative to stimulus intensity. Each point represents the average latency of R_1 (squares) and R_2 (circles) responses emitted by 20 Ss.

Fig. 4. Comparison of response probability functions for Ss who made many errors during prior discrimination training. The unfilled symbols represent response probabilities for Group I and the filled symbols represent response probabilities for Group II with an average of 3.6 and 15.7 errors, respectively, during the last half of discrimination training.

Fig. 5. Mean latency functions (R_1 and R_2 combined) for Ss who made few errors (I) and Ss who made many errors (II) during prior discrimination training

Fig. 6. Relations among response probability, latency, and topography in stimulus generalization.

Top: Median frequency in cps (circles) and the range (vertical lines) of high and low pitch responses as a function of stimulus intensity for each of 3 Ss. The dashed horizontal lines represent the vocal pitches previously differentiated.

Bottom: Response probabilities (squares) equal the ratio of the number of low pitch responses emitted to the number of stimulus presentations (10) at each intensity. The hexagons represent the average latency of high and low pitch responses to each stimulus intensity.

Fig. 7. Spectrographic patterns which were converted to sound by the Pattern Playback to form the speech stimuli of the experiment (after Liberman _et al_., 1961).

Fig. 8. Conditional probability and average latency of vocal responses to synthetic speech stimuli under three sets of instructions (see text).

Fig. 9. The average latencies of the vocal responses (Fig. 8) have been subdivided in terms of their R_1 and R_2 components.

Fig. 10. "Psychometric curves for relative intensity judgments" (after Kellogg, 1931)

Left side: R_1 ("left-side-is-darker") and R_2 ("right-side-is-darker") are the response probability functions and L_T is the average latency curve for 5 Ss judging seven stimulus intensity pairs for a total of 240 judgments on each pair.

Right side: L_{R_1} and L_{R_2} are the mean latency functions for the R_1 and R_2 response respectively.

The stimulus designations refer to the half of the visual field which had the lower luminance. The stimulus continuum ranges from L_3 (lowest luminance of left field) through E (luminance of the two fields equal) to R_3 (lowest luminance of right field).

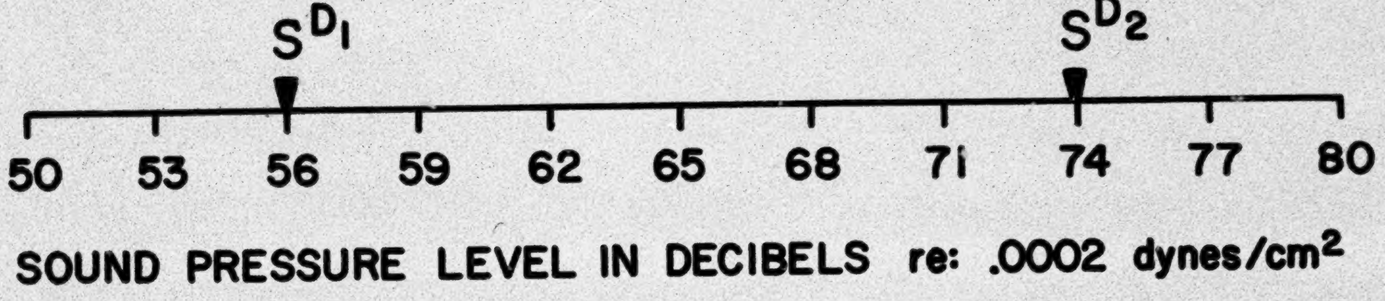
S^{D_1}
S^{D_2}
SOUND PRESSURE LEVEL IN DECIBELS re: .0002 dynes/cm^2

Fig. 1

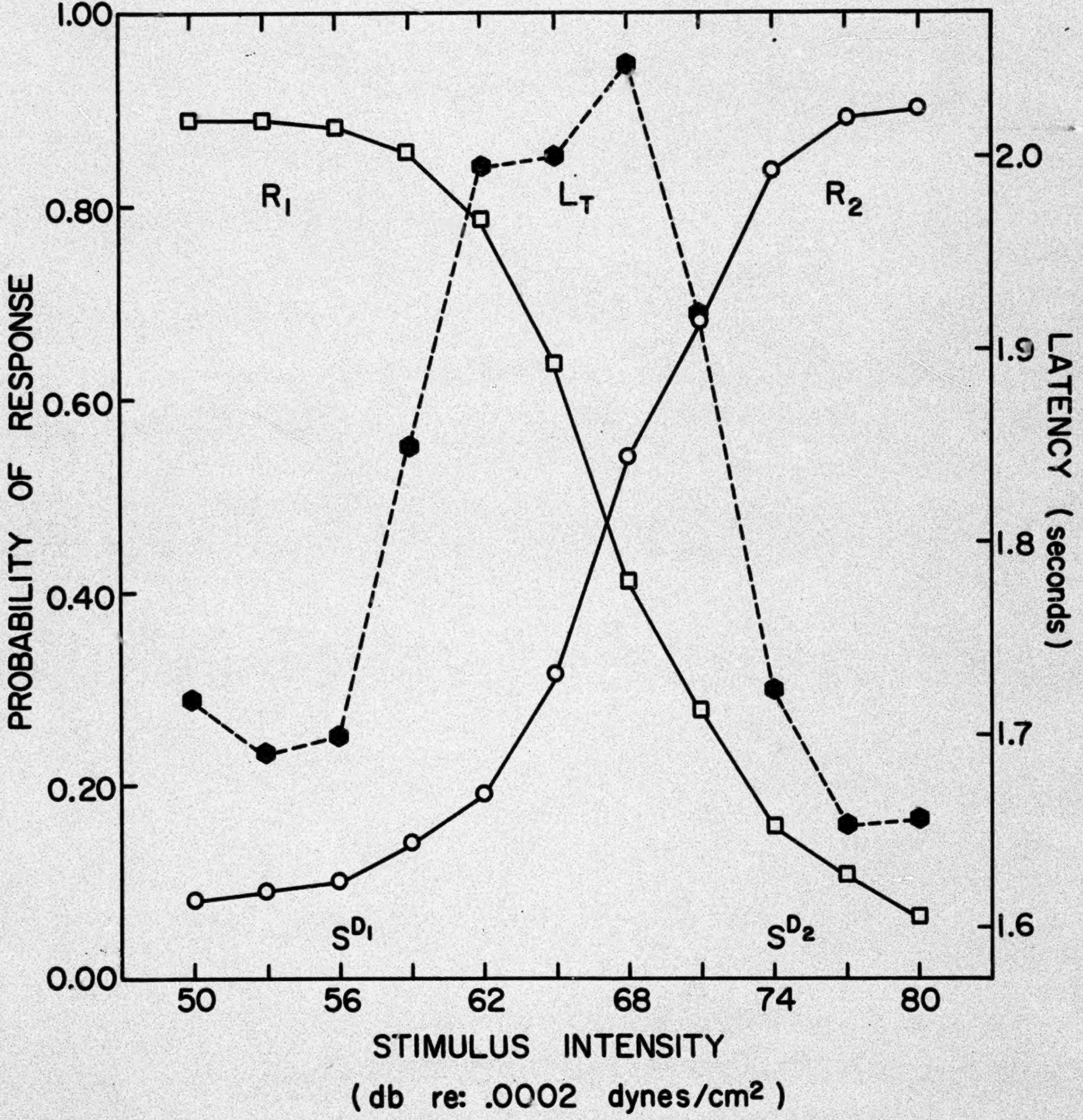
PROBABILITY OF RESPONSE
1.00
0.80
0.60
0.40
0.20
0.00
R1
LT
R2
SD1
SD2
LATENCY (seconds)
2.0
1.9
1.8
1.7
1.6
50
56
62
68
74
80
STIMULUS INTENSITY
(db re: .0002 dynes/cm2)

Fig. 2

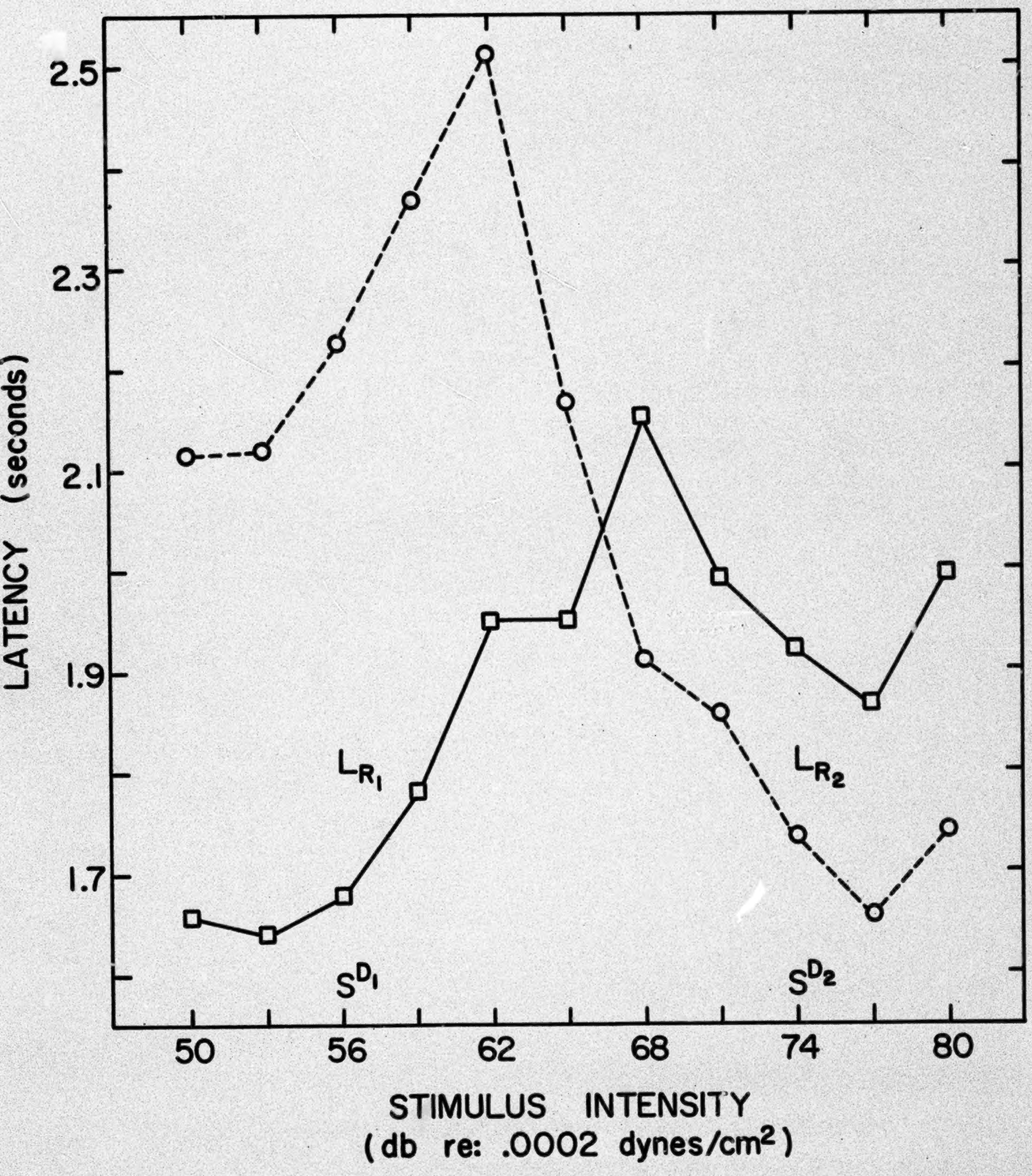
LATENCY (seconds)
2.5
2.3
2.1
1.9
1.7
L_{R_1}
L_{R_2}
S^{D_1}
S^{D_2}
50
56
62
68
74
80
STIMULUS INTENSITY
(db re: .0002 dynes/cm^2)

Fig. 3

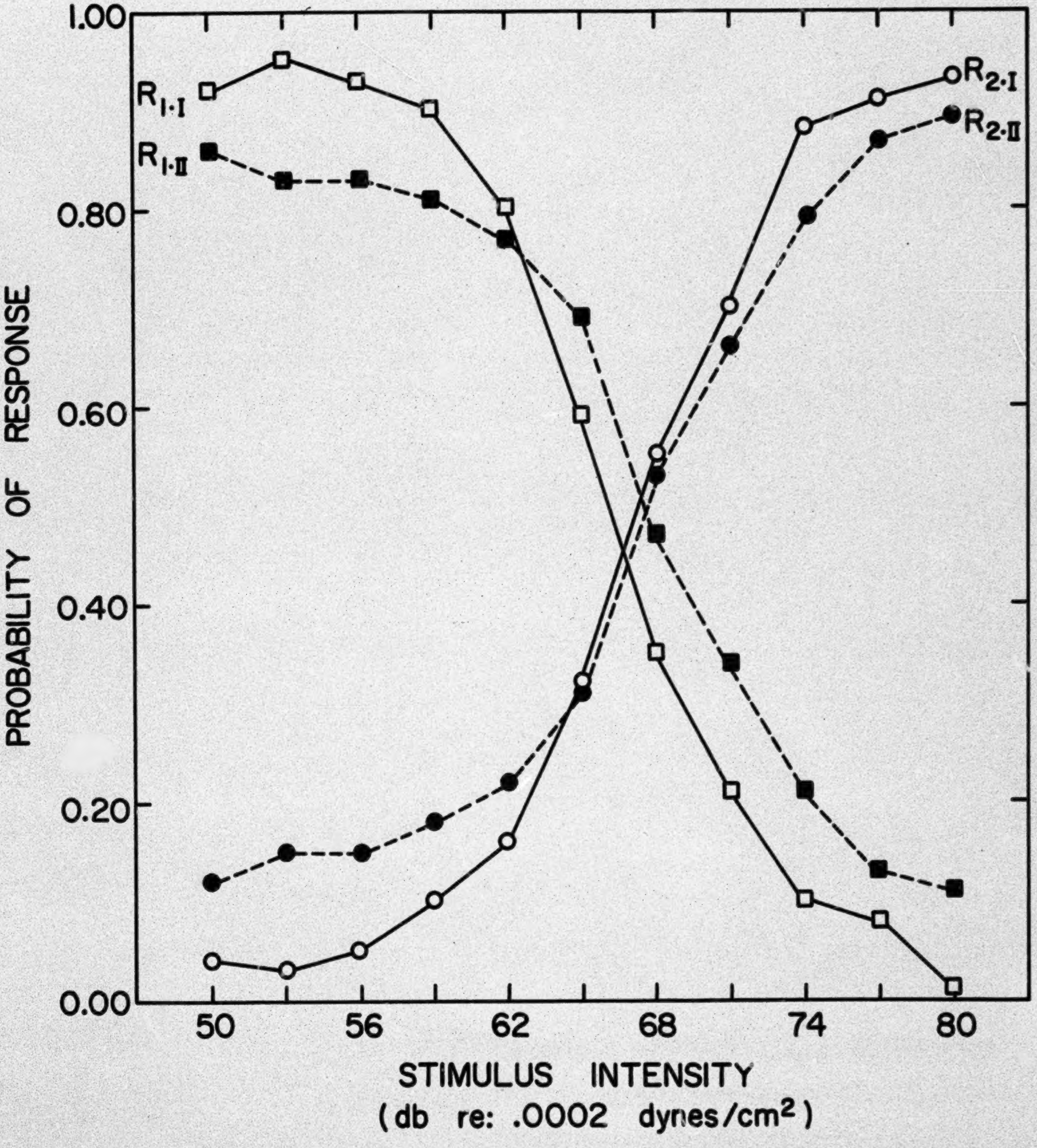
1.00
0.80
0.60
0.40
0.20
0.00
PROBABILITY OF RESPONSE
50
56
62
68
74
80
STIMULUS INTENSITY
(db re: .0002 dynes/cm²)
$R_{1 \cdot I}$
$R_{1 \cdot II}$
$R_{2 \cdot I}$
$R_{2 \cdot II}$

Fig. 4

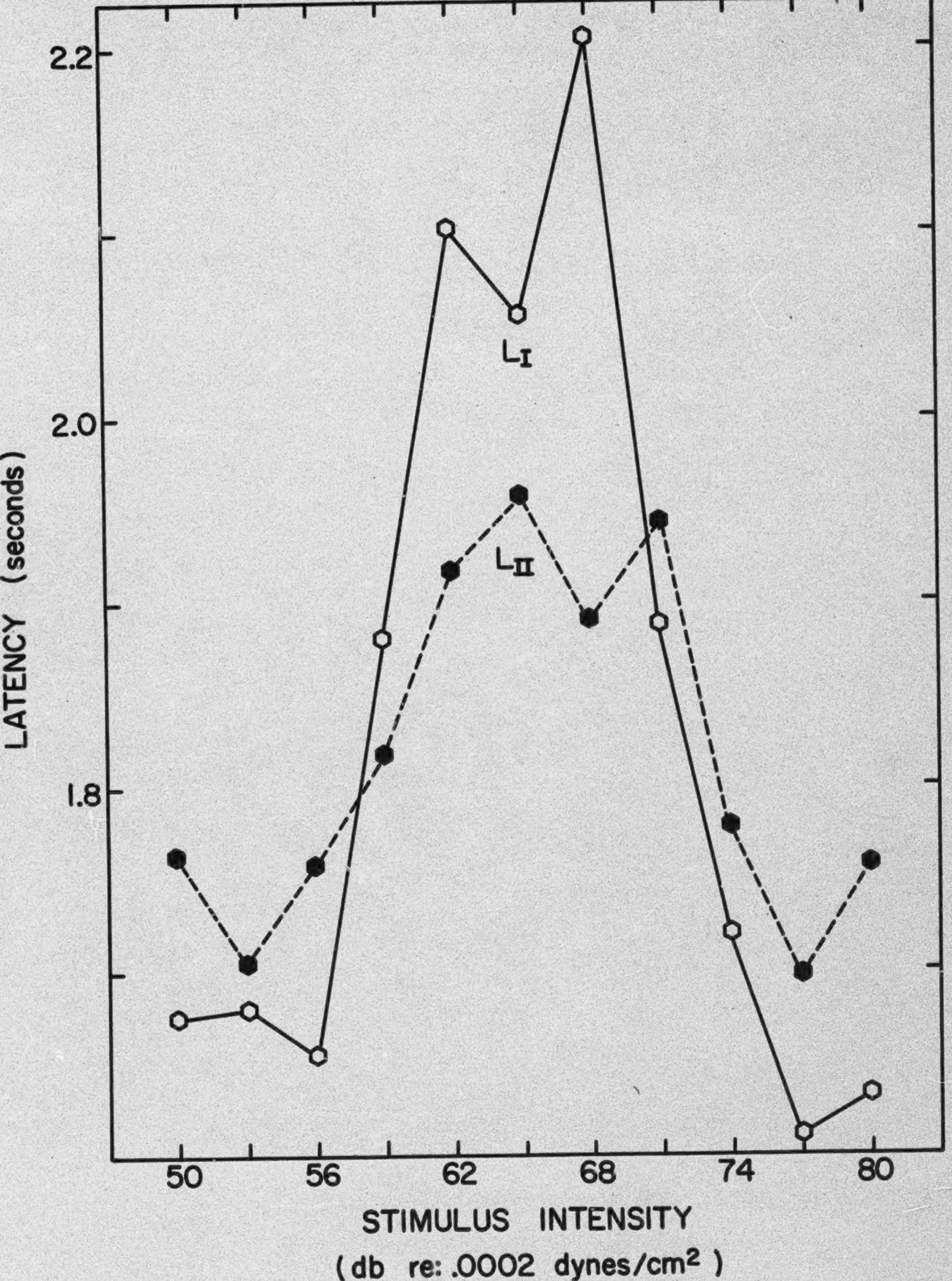
2.2
2.0
1.8
LATENCY (seconds)
L_I
L_{II}
50
56
62
68
74
80
STIMULUS INTENSITY
(db re: .0002 dynes/cm²)

Fig. 5

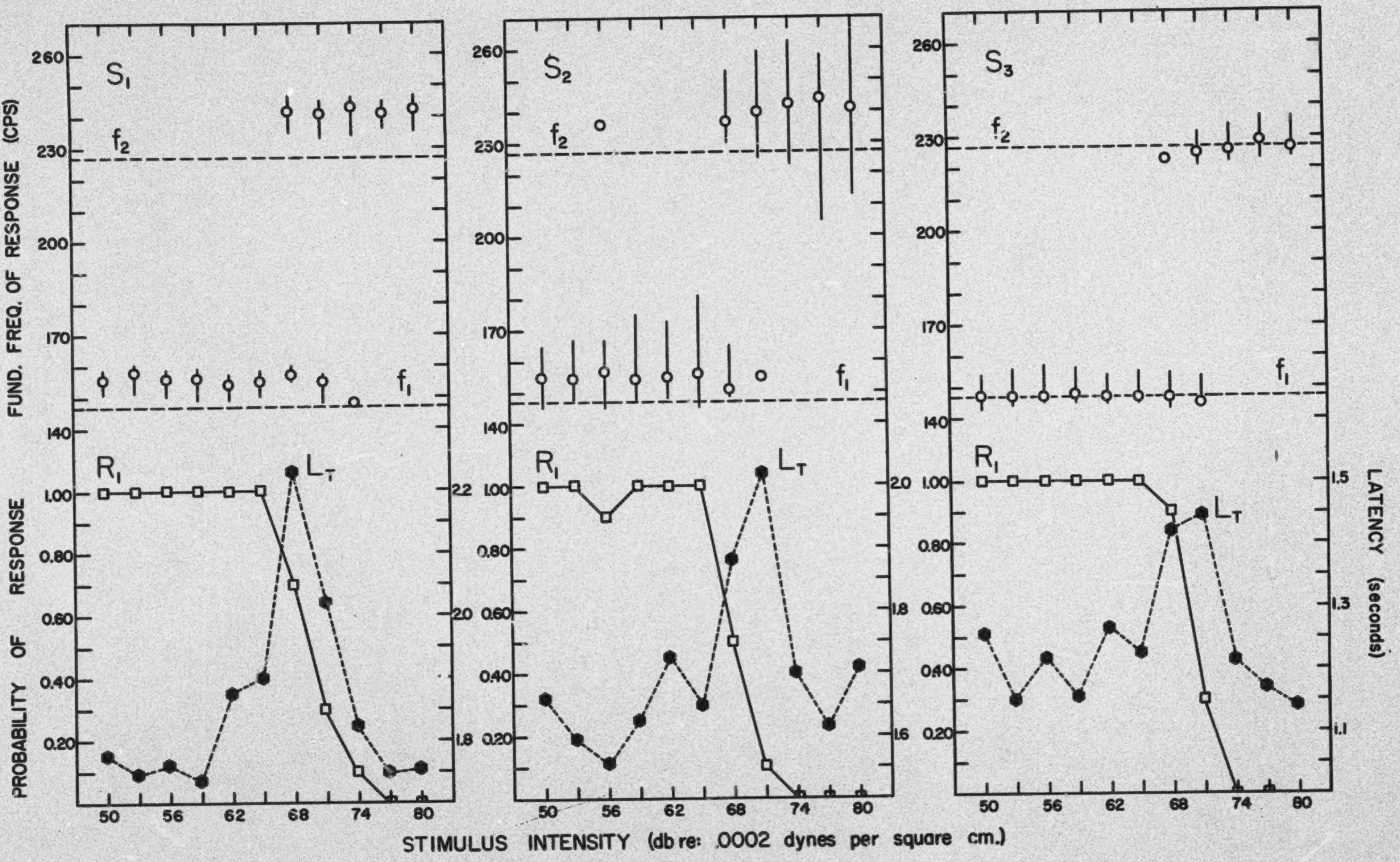
S₁
S₂
S₃
f₂
f₁
R₁
L_T
FUND. FREQ. OF RESPONSE (CPS)
PROBABILITY OF RESPONSE
LATENCY (seconds)
260
230
200
170
140
1.00
0.80
0.60
0.40
0.20
2.2
2.0
1.8
1.6
1.5
1.3
1.1
50
56
62
68
74
80
STIMULUS INTENSITY (db re: .0002 dynes per square cm.)

Fig. 6

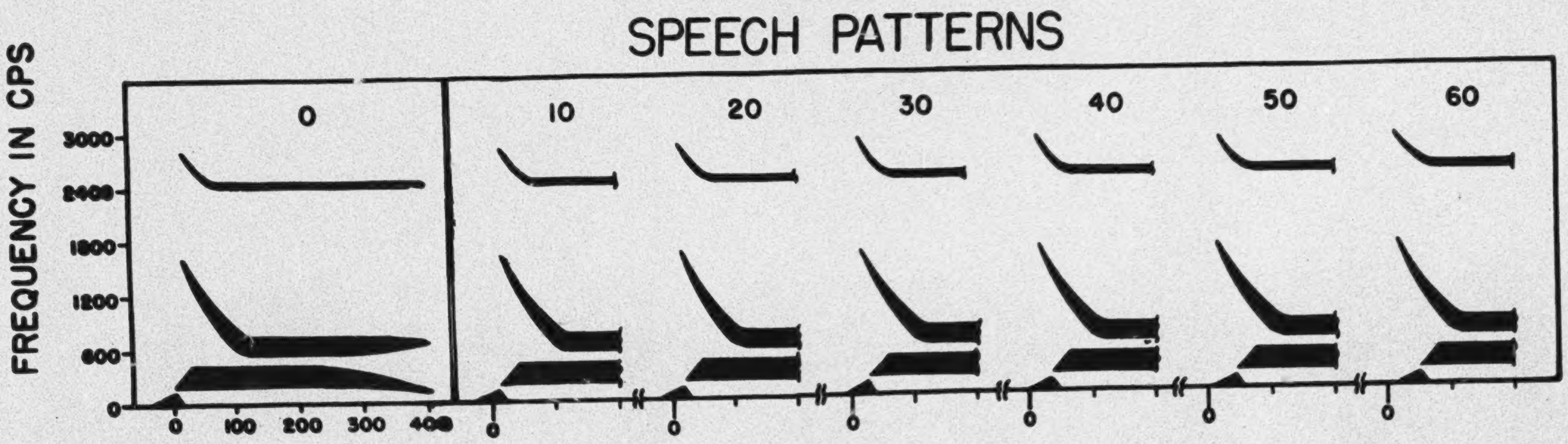
SPEECH PATTERNS
FREQUENCY IN CPS
3000
2400
1800
1200
600
0
0
100
200
300
400
0
10
20
30
40
50
60

Fig. 7

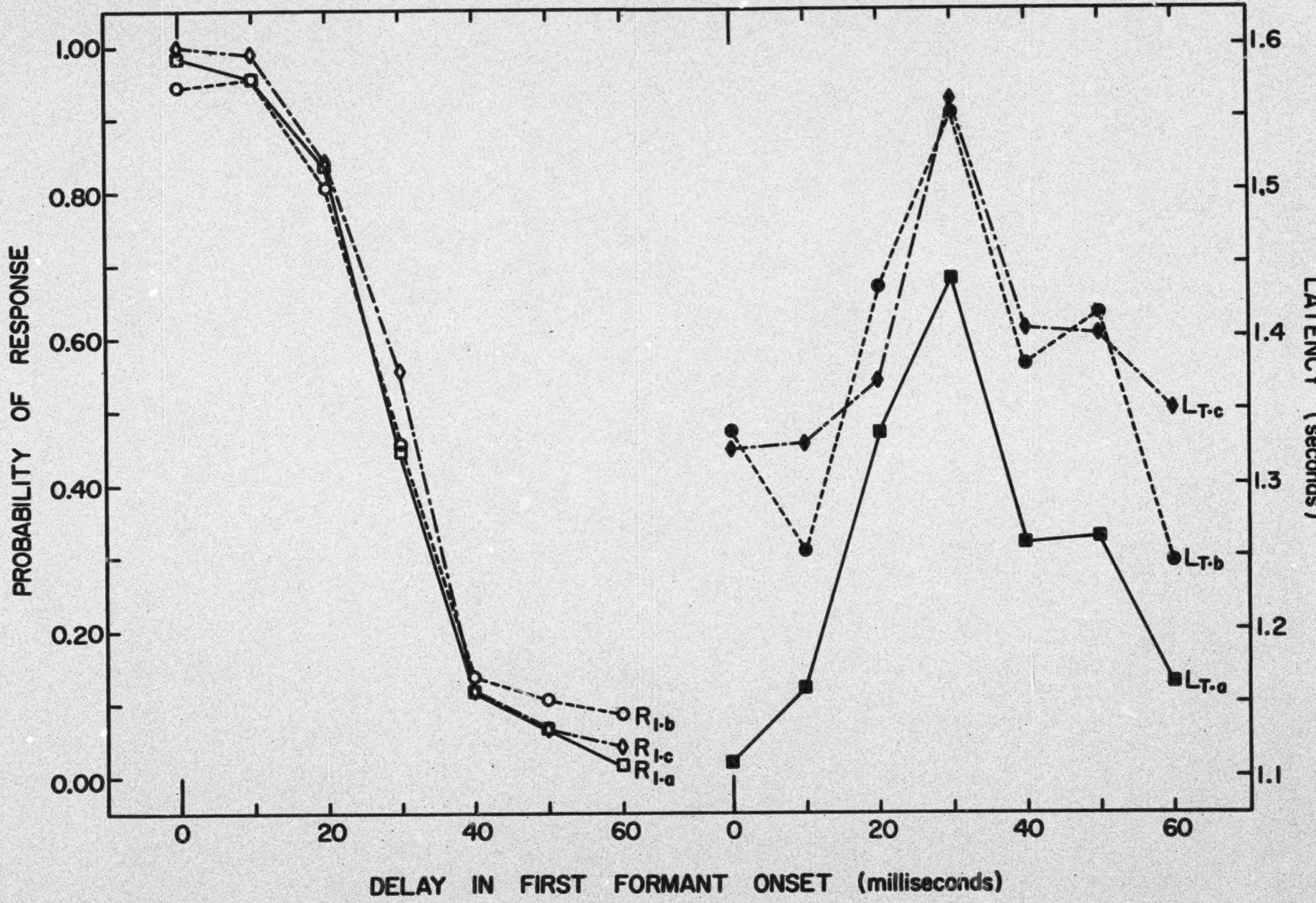
PROBABILITY OF RESPONSE
1.00
0.80
0.60
0.40
0.20
0.00
LATENCY (seconds)
1.6
1.5
1.4
1.3
1.2
1.1
0
20
40
60
0
20
40
60
$R_{1\cdot b}$
$R_{1\cdot c}$
$R_{1\cdot a}$
$L_{T\cdot c}$
$L_{T\cdot b}$
$L_{T\cdot a}$
DELAY IN FIRST FORMANT ONSET (milliseconds)

Fig. 8

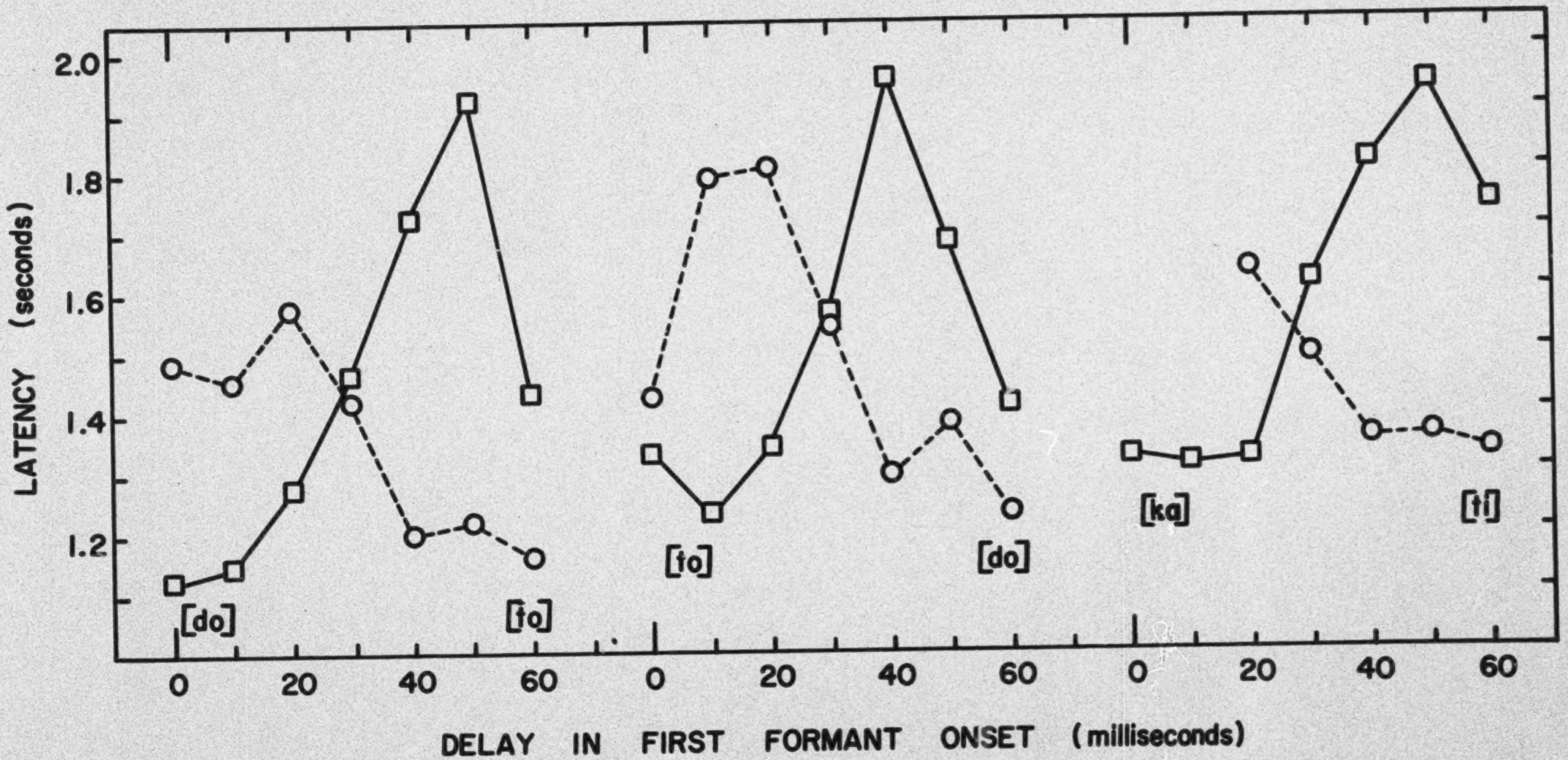
2.0
1.8
1.6
1.4
1.2
LATENCY (seconds)
[do]
[to]
[to]
[do]
[ka]
[ti]
0
20
40
60
0
20
40
60
0
20
40
60
DELAY IN FIRST FORMANT ONSET (milliseconds)

Fig. 9

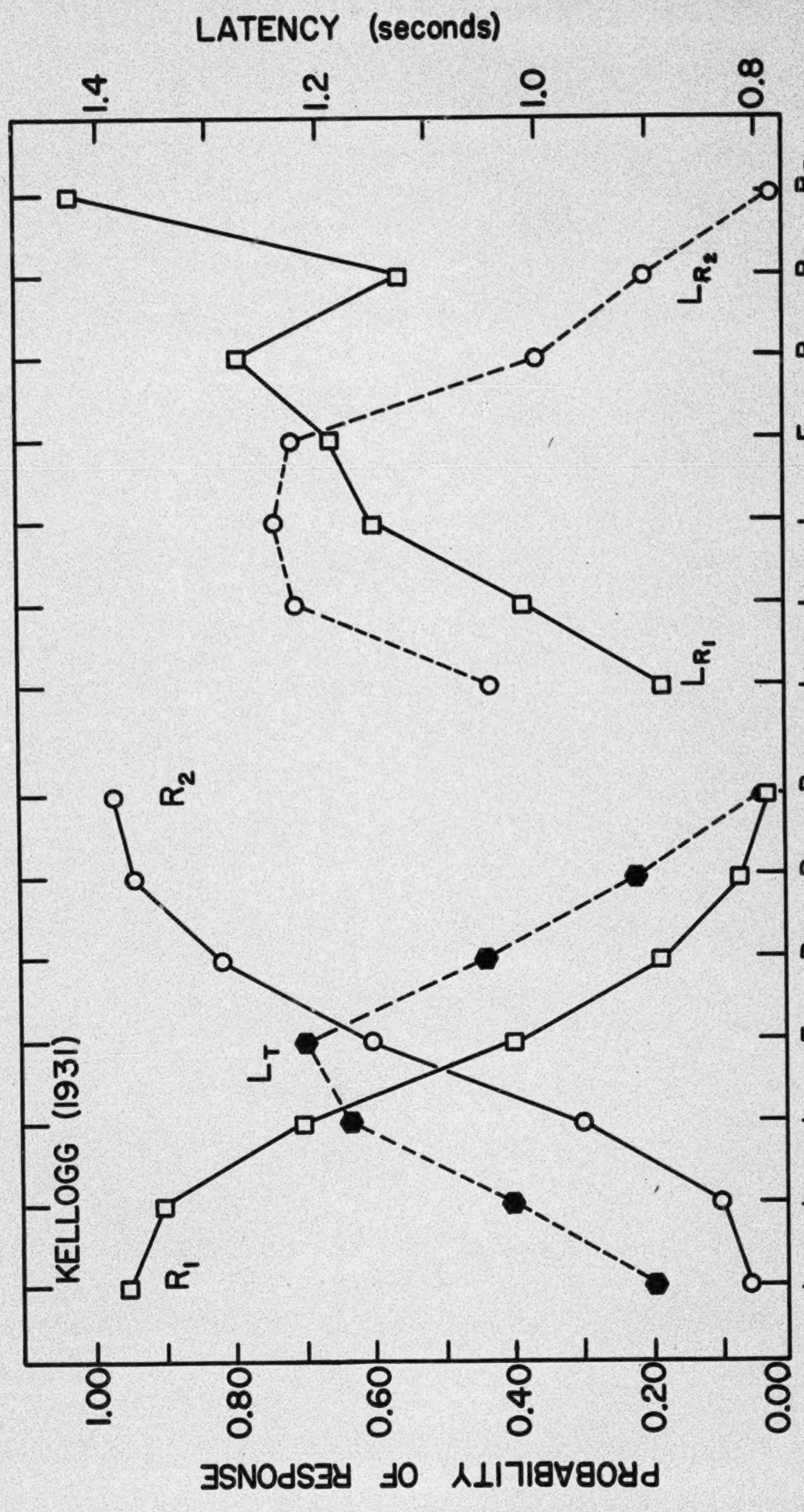
KELLOGG (1931)
LATENCY (seconds)
1.4
1.2
1.0
0.8
PROBABILITY OF RESPONSE
1.00
0.80
0.60
0.40
0.20
0.00
STIMULUS PAIR
L3
L2
L1
E
R1
R2
R3
R_1
R_2
L_T
L_{R_1}
L_{R_2}

Fig. 10

THE EFFECTS OF CHANGING VOWEL PARAMETERS ON PERCEIVED LOUDNESS AND STRESS.

I: DOES AUTOPHONIC LEVEL AFFECT THE LOUDNESS FUNCTION?

Harlan Lane
The University of Michigan

THE EFFECTS OF CHANGING VOWEL PARAMETERS ON PERCEIVED LOUDNESS AND STRESS. I: DOES AUTOPHONIC LEVEL AFFECT THE LOUDNESS FUNCTION?

Harlan Lane
Communication Sciences Laboratory
The University of Michigan

"Speech is perceived by reference to articulation." This hypothesis was offered by Liberman in 1957, after summarizing research at the Haskins Laboratories on the cues relevant to the recognition of speech sounds. The hypothesis was based upon the observation that different acoustic stimuli give rise to the same phonemic identifications and that speech sounds are discriminated largely to the extent that they can be identified as belonging to different phoneme categories. Liberman suggested that certain previously conditioned articulatory responses and their consequent sensory effects "mediate between the acoustic stimulus and the event we call perception." Thus, different acoustic stimuli come to sound alike to the extent that they are produced by the same gross articulatory movements. This "mediation-hypothesis" has received substantial support in the recent findings of psychoacoustic and electromyographic research. In a report of their findings of phonemic contrast induced by silence, Bastian *et al*. (1961) concluded: "...the categorical perception of the consonants may be explicable in terms of the categorized nature of their articulatory gestures."

Additional support for the mediation hypothesis comes from a second quarter: investigations of speech loudness and stress. In 1932 S. Jones wrote "Accent is *sui generis*, depending for its perception on the kinaesthetic sense. The listener refers what he hears to how he would say it. Thus he translates exteroceptor into proprioceptor sensations, the kinaesthetic memory serving as stimulus." Following an investigation of the action of the respiratory muscles during speech, Draper *et al*. (1952) concluded that "naive listeners, obeying an instruction to consider the loudness of sounds in continuous speech, do not assess the acoustic properties of the sounds but consider, instead, the pressure which would be required below the vocal cords." With the same point of departure, Ladefoged (1958) has written: "Statements about stress are usually best regarded as statements about the speaker's muscular behavior (or about the action of the listener's muscles which would have to be made in order to produce similar sounds)." Lehiste and Peterson (1959) have shown that when two vowels, generated with unequal effort, are presented at the same sound pressure level, the listener identifies the vowel produced with greater effort as louder. This finding is seen by the authors to support their contention that "the listener interprets speech according to the properties of the speech production mechanism rather than according to the psychophysical principles of the perception of abstract sounds."

One test of the mediation hypothesis as it is applied to the perception of vowel loudness, would be to obtain the loudness functions for speech from speakers and listeners and see if these are identical or, at least, similar. Lane, Catania and Stevens (1961) have shown that they are not. The speaker's numerical estimation of his own vocal level, *the autophonic response*, grows as as the 1.1 power of the actual sound pressure produced, whereas a listener's estimates of the same productions grow as the 0.7 power of the sound pressure.

The disparity between these functions suggests that (1) the speaker does not rely solely upon his perception of loudness in judging his autophonic level and (2) the listener does not rely solely upon his perception of autophonic level in judging loudness.

This conclusion does not lend support to the mediation hypothesis in its "strong" form. Since the acoustic parameters of a vowel are related to the autophonic level at which it is generated (Fant, 1958) a weaker form of the hypothesis might be: when the cues to effort are present, loudness grows more rapidly as a function of sound pressure level than when they are not. This hypothesis may also be rejected. Lane, Catania and Stevens (1961) have shown that the loudness-of-speech function is no different when listeners judge different playback levels of a single /a/, recorded at moderate level, or when they judge live vocal productions in which the speaker varies his voice over a 30 db range. "Under such wide changes in voice level, the quality of the sound inevitably alters, but this fact did not alter the measured exponent [of the loudness function]."

These findings, obtained with ratio-scaling techniques, show no evidence for the redintegration to which Jones, Draper and Ladefoged refer, and raise some question as to the validity of the mediation hypothesis, at least in the context of vowel loudness judgments.

While the rate of change of vowel loudness as a function of sound pressure level does not seem to be affected by judgments of effort or autophonic level, it may be that loudness is thus affected. That is, vowel parameters other than sound pressure may influence the y-intercept rather than the slope of the loudness function. A subsequent study in this series will examine this possibility. The present study undertakes a more extensive examination of the relation between autophonic level and the rate of change of loudness as a function of sound pressure level.

METHOD

There were ten stimulus series, each of which comprised 28 speech or nonspeech stimuli recorded and later reproduced (at 7.5 ips) with professional tape recorders (Ampex 300 and 350).

Series (1). Autophonic level and sound pressure level covaried. The experimenter produced the phoneme /a/ for approximately two seconds at seven intensity levels (read on a Ballantine rms vt voltmeter) spaced equally over a 40 db range. Spectrograms were prepared (Western Electric BTL-2 spectrograph) from a tape recording of the vocal responses. The average speech power of the vowel (integrated over 10 milliseconds) was displayed as a function of time on an oscilloscope (Tektronics 533) and photographed. The entire series of seven productions was repeated four times, until the two analyses described showed that (a) the fundamental frequency did not vary by more than two cps and the average speech power by more than 0.5 db during the "steady state" of each vowel and (b) the decibel difference between the peak average speech power of successive productions was 5 db ± 0.5 db. Samples of each vowel were then obtained with an electronic switch (Grason-Stadler 829S119) controlled by an interval timer (Grason-Stadler 471) so that only a central portion of the vowel would be selected; the duration of each sample was 500 msec. and the rise time

100 msec. Using magnetic recording techniques, a series of 28 stimuli was prepared in which each of the seven samples was presented four times in irregular order. When copying from one tape recorder (Ampex 300) to another (Ampex 350) a bandpass filter (Krohn-Hite 310-AB, set at 100-4,000 cps) was interposed. This improved the signal to noise ratio on the tape recordings which always exceeded 30 db. Table I shows the acoustic parameters of the seven productions of the vowel /a/ that were employed.

Series (2), (3), and (4). Autophonic level constant, SPL varied. Each of these stimulus series consisted of a single /a/ (numbers 5, 15, and 25, respectively, in Table I) presented at several intensities. To prepare the series, each of the three stimuli was recorded first on a separate magnetic tape loop. The recording was then played back repeatedly and the signal sent through a calibrated attenuator (Hewlett Packard 350A) to a second continuous tape recording. The attenuator was adjusted during the four second interstimulus interval so that the recorded stimulus series consisted of 28 presentations of the stimulus, four at each of seven sound pressure levels, equally spaced over a 30 db range. The only within-series variable was the sound pressure level; the only between-series variable was the acoustic parameters of the vowel employed.

Series (5). Synthesized /a/ (pulse spectrum). An oscillator (Hewlett-Packard 207A) generated a sinusoid at 125 cps (calibrated with a Hewlett-Packard 522B electronic counter). This signal drove a pulse generator (SRL) which applied 125 cps and its harmonics to three filters in series (two Krohn-Hite 310-AB and a Dytronics 720). The output of the last filter was recorded on a four-channel tape recorder (Ampex 300-4). To obtain the first formant of the synthetic vowel, the three filters were set at 760 cps and their output was recorded for five minutes at -5VU. With the three filters in cascade the half-power bandwidth was 91 cps and the attenuation 60 db/octave (calibrations with a General Radio Sweep frequency oscillator 1304-B and graphic level recorder, model 1521-A). The filter settings were then changed to 1065 cps (bandwith 128 cps) and the second formant recorded adjacent to the first on a second tape track.

The two concurrent formants were played back and combined at equal rms voltages with an electronic mixer (SRL). An electronic switch (rise time 100 msec.) sampled 500 msec. portions of the continuous two-formant signal every four seconds. The synthetic /a/ was then sent through an attenuater to a tape recorder. During the inter-stimulus interval, the attenuator was adjusted so that the recorded series of intensities was identical with that in series (1).

Each of the vowel parameters chosen only approximate those observed with live productions of the phoneme /a/. This simplification serves the present purpose which is to compare the loudness function for an idealized vowel, with relatively low intelligibility, to that for the more complex "natural" signal.

Series (6). Synthesized /a/ (noise spectrum). The pulse generator in the apparatus arranged for series (5) was replaced by an equal-amplitude random noise generator (Grason-Stadler 455B) and an identical procedure was followed. The intensity series of synthetic /a/'s obtained simulated the unvoiced or whispered phoneme.

Series (7) Synthesized formant, SPL constant, pitch varied. The same arrangement of equipment used to generate the first formant in series (5) was employed. However, the amplitude of the formant remained constant; the fundamental frequency was varied over the range 100-220 cps. Seven "pitch levels" at 20 cycle intervals were each presented four times in irregular order. This series was designed to serve as a non-speech control for series (8).

Series (8) SPL constant, autophonic level varied. Stimulus series (1) was copied from one tape recorder to a second with a high-fidelity amplifier (SRL) and attenuator interposed. Compensatory adjustments in amplification were made during the inter-stimulus intervals so that all stimuli were at the same intensity on the final recording (± 0.25 db); this was verified by processing the tape recording with an average speech power circuit and a recording oscillograph (Minneapolis-Honeywell Visicorder).

Series (9) Seven autophonic levels, generated at five db intervals, reproduced at two db intervals. Following a procedure simular to that employed in series (8), a tape recording was prepared by adjusting intensity differences among the seven autophonic levels of series (1), so that there were two-decibel differences in intensity between successive stimuli. As in all other series, each stimulus appeared four times, for a total of 28 stimulus presentations.

Series (10) 1,000 cps tone, SPL varied. A tape recording was prepared whose format was identical to that in series (1). Twenty-eight samples of a 1,000 cps tone (500 msec. duration, 100 msec. rise time), four at each of seven intensity levels, were recorded in irregular order. This series served as a control for the particular adaptation of the method of magnitude estimation and the stimulus recording and reproduction techniques employed in series (1) through (9).

Each of ten male undergraduates, none of whom had participated previously in psychophysical research, served in sessions lasting approximately 30 minutes. The subject was seated in an anechoic chamber in front of a microphone. He wore a binaural headset with matching PDR-10 earphones and MX41-AR cushions. The frequency calibration obtained for one of the earphones is shown in Fig. 1. The frequency response of the second earphone did not differ from that shown by more than 3 db at any frequency. The experimenter, located in an adjacent room, presented the stimulus series in a different irregular order to each subject, except that series (10) was always presented first. The first stimulus in each series served as the modulus or standard (Stevens, 1958). It had a median sound pressure level and/or fundamental frequency for that particular series.

The output of the tape recorder that presented the stimuli was filtered (bandpass, 100 to 4,000 cps), amplified by a transistorized amplifier with low signal to noise ratio and flat frequency response (± 1 db 20-20,000 cps) and applied to the headphones. A 1,000 cps tone at 30 db below 0 VU had a sound pressure level, when transduced by the earphone, of 50 decibels. As a consequence, the range of intensity levels presented in series (10) was exactly 50-80 db (SPL). This range can only be given approximately for the nine other series because they involved signals with complex acoustic spectra: series (1) through (6), 50-80 db; series (2) and (8), constant sound pressure level, 80 db; series (9) 68-80 db.

These instructions were read to the subject:

> "This is an experiment to see how you perceive the loudness of some sounds. Your task will be to give a numerical estimate of the loudness of each stimulus as it comes along.

The first sound on each tape is the standard loudness which you are to call "ten." When you hear it, say, "ten." Then, after you hear the second sound tell me the numerical value of its loudness and so on through the tape. Always assign numbers to the stimuli in the same proportion to "ten" as their loudness is to the standard. For example if the second stimulus is twice as loud as the standard call it twenty, if the second stimulus is half as loud as the standard call it ...(five). Are there any questions? Let's begin. Remember, the first stimulus is the standard and you are to call it 'ten'."

RESULTS AND DISCUSSION

Table II shows the median loudness estimates obtained in each stimulus series. In every series except (7), the growth of loudness as a function of the stimulus variable is well represented by a straight line in a log-log plot. This finding is further evidence for the general contention that a power law describes the operating characteristics of sensory transducers (Stevens, 1957, 1960).

To a first decimal approximation, the exponent of the power function governing vowel loudness (see column D, Table II) is the same in stimulus series (1) through (5). The autophonic level at which the vowel was generated did not affect the measured exponent of 0.4. Lane, Catania and Stevens (1961) have also shown that the slope of the vowel loudness function is independent of autophonic level. However, these authors found that vowel loudness grows as the 0.7 power of the sound pressure level. There are few procedural differences between the two studies. One salient difference is that the present study employed untrained subjects, whereas the former employed graduate students in psychology who were trained psychophysical observers. The use of untrained subjects and hence greater variability in numerical estimates are usually associated with a flattening of the loudness function (Stevens and Poulton, 1956; Stevens and Tulving, 1957). The vowel loudness function was found to have a slope of 0.4 on five occasions in the present study. Pollack (1952) has shown that the loudness of a tape recorded passage of spoken text grows as the 0.4 power of the average sound pressure level over the range 50 to 80 db. Based on the available evidence, 0.4 may be the best estimate of the slope (exponent) of the vowel loudness function.

Table II shows that the power law governing the loudness of a two-formant synthesized /a/ with noise spectrum [Series (6)] and that for a 1,000 cps tone [Series (10)] have the same exponent, 0.5. These findings confirm those of earlier research. J. C. Stevens and E. Tulving (1957) have shown that median estimates of the loudness of white noise, given by 70 untrained observers, grow as the 0.5 power of the sound pressure over the range 55-105 db. S. S. Stevens and E. C. Poulton (1956) found essentially the same results with unpracticed observers estimating the loudness of a 1,000 cps tone. In the light of these findings, it may be inferred that the stimuli of series (6), generated with an equal excitation source, were not perceived as vowels at all but rather as bands of noise. This seems the more likely when we recall that series (6) was presented at sound pressure levels much greater than those normally associated with a whispered vowel. All of the stimuli in the present study were presented at sound pressure levels of 50 db or greater in order to reduce three kinds of confounding effects that take place at low intensities: (1) Stevens (1961) has summarized the evidence for a "mid level bulge" in the power function relating the loudness of complex sounds to their sound pressure level. (2) Scharf (1959) has shown that the opposite effect may take place at low sound pressure levels. (3) Ratio scales of subjective magnitudes on several sensory continua have been shown to turn concave downward near threshold. (Summarized by Stevens, 1960).

In series (8) seven autophonic levels were adjusted in intensity so that all stimuli were of equal sound pressure level (80 db). When the median loudness estimates (transformed to db) were plotted as a function of the decibel differences in sound pressure among the *original* autophonic levels (before processing) a straight line with slope 0.1 was obtained. It appears, therefore, that vowel parameters other than intensity, which are correlated with autophonic level, may influence loudness judgment. The opposite conclusion was reached earlier in this study based on the findings for series (1) through (6). The apparent contradiction is resolved by underscoring a difference between series (8) and the others. In this series, untrained subjects were instructed to estimate the relative loudness of stimuli *that did not differ in sound pressure level*. Given instructions that implicitly required changing numerical estimates, most of the subjects were influenced by the spectrum of the signal in the absence of the stimulus changes that normally control loudness judgments. (Only one subject assigned the value of the modulus, "ten," to all stimuli.) This interpretation is supported by the outcome of series (7), in which a band of harmonics with changing fundamental frequency was presented at constant sound pressure level. Once again, untrained subjects were constrained to give loudness estimates in the absence of changes in sound pressure level; as a consequence, the estimates were influenced slightly by changes in the frequency spectrum of the stimulus.

It is not clear whether the effect of spectrum on loudness estimates observed in series (7) and (8) is due to one or more of the following: the correlation between vocal pitch and vocal sound pressure level; increasing sensitivity of the ear with increasing pitch in the frequency range employed; a purely intraverbal linkage between high pitch - high intensity and low pitch - low intensity evoked by all auditory stimuli [cf. the discussion of vowel symbolism by Brown (1958)]. In any event it is clear that the effect of autophonic level on loudness takes place only when the typical relations among sound pressure and other vowel parameters are severely distorted. The procedures which would be most sensitive to the effects of this distortion are those requiring ordinal judgments from the subject. If two vowels of identical sound pressure level are presented and S is instructed to choose the louder, any difference whatsoever between the two vowels might be seen as influencing loudness judgments. These ordinal data can yield a misleading image of the magnitude of the effect; ratio-scaling techniques show it to be quite small.

The findings obtained with series (9) show the effect on loudness when an intermediate amount of distortion is introduced in the relation between sound pressure level and autophonic level, operating in concert. The slope of the loudness function is also intermediate. Judging from slope alone, sound pressure is the dominant factor.

In general, it is true of perceptual phenomena that the sum of the separate effects of cues operating in isolation does not equal their combined effect when operating in concert. Autophonic level seems to affect loudness when sound pressure does not vary [series (8)]. Sound pressure controls loudness judgments when autophonic level is held constant [series (2), (3), (4)]. When both sound pressure and autophonic level covary [series (1), (9)], the effects do not combine, but, rather, sound pressure emerges as the controlling variable. This finding reveals the weakness of a pons asinorum for the student of perception. By isolating the components of a complex discriminative stimulus, the magnitude of their separate effects on the perceptual response may be assessed. However, when these components operate in concert, their relative weights may be radically different.

SUMMARY

The hypothesis that speech is perceived by reference to articulation was examined in the context of vowel loudness judgments. The normal relations among autophonic level and sound pressure level were variously distorted in the preparation of ten stimulus series. Ten untrained observers gave numerical estimates of the loudness of stimuli in each series.

When sound pressure cues to loudness were distorted and the subject was required to give loudness estimates, autophonic level had a demonstrable effect on these judgments. However, distortion of autophonic cues to loudness did not influence the vowel loudness function, which was found to grow as the 0.4 power of the sound pressure level.

FOOTNOTE

1. The assistance of Mr. D. R. Brinkman is gratefully acknowledged. This research was performed pursuant to a contract with the United States Office of Education, Language Development Section.

REFERENCES

Bastian, J., Eimas, P. D., and Liberman, A. M. Identification and discrimination of a phonemic contrast induced by silent interval. (Paper read at the 61st Meeting of the Acoustical Society of America) *J. Acoust. Soc. Am.*, 33, 1961, 842.

Brown, R. *Words and Things*. Glencoe, Illinois: Glencoe Freepress, 1958.

Draper, M. H., Ladefoged, P. and Whitteridge, D. Respiratory muscles in speech. *J. Sp. Hear. Res.*, 2, 1952, 16-27.

Fant, C. G. M. Modern instruments and methods for acoustic studies of speech. In *Proc. VIII Int. Congr. of Linguists*. Oslo: Oslo University Press, 1958.

Jones, S. The accent in French --- what is accent? *Le Maitre Phonetique*, 40, 1932, 74-75.

Ladefoged, P. Syllables and stress. *Miscellanea Phonetica*, 3, 1958, 1-14.

Lane, H. L., Catania, A. C., and Stevens, S. S. Voice Level: autophonic scale, perceived loudness, and effects of sidetone. *J. Acoust. Soc. Am.*, 33, 1961, 160-177.

Lehiste, I. and Peterson, G. E. Transitions, glides, and dipthongs. *J. Acoust. Soc. Am.*, 33, 1961, 268-277.

Lehiste, I. and Peterson, G. E. Vowel amplitude and phonemic stress in American English. *J. Acoust. Soc. Am.*, 31, 1959, 428-435.

Liberman, A. M., Harris, K., Kinney, J. A., and Lane, H. L. The discrimination of relative onset time of the components of certain speech and non-speech patterns. *J. exp. Psychol.*, 61, 5, 1961, 379-389.

Liberman, A. M. Some results of research on speech perception. *J. Acoust. Soc. Am.*, 29, 1957, 117-123.

Peterson, G. E. Parameters of vowel quality. *J. Sp. Hear. Res.*, 4, 1961, 10-29.

Pollack, I. On the measurement of the loudness of speech. *J. Acoust. Soc. Am.*, 24, 1952, 323.

REFERENCES (Concluded)

Scharf, B. Critical bands and the loudness of complex sounds near threshold. J. Acoust. Soc. Am., 21, 1959, 365-370.

Stevens, J. C. and Tulving, E. Estimation of loudness by a group of untrained observers. Amer. J. Psychol., 70, 1957, 600-605.

Stevens, S. S. and Poulton, E. C. The estimation of loudness by unpracticed observers. J. exp. Psychol., 51, 1956, 71-78.

Stevens, S. S. On the psychophysical law. Psychol. Rev., 64, 1957, 153-181.

Stevens, S. S. Problems and methods of psychophysics. Psychol. Bull., 54, 1958, 177-196.

Stevens, S. S. Procedure for calculating loudness. Mark VI. Psycho-Acoustic Laboratory, Harvard University, Report PNR-253, 1 March 1961.

Stevens, S. S. The psychophysics of sensory function. Amer. Scient., 48, 1960, 226-253.

TABLE I

ACOUSTIC PARAMETERS OF SEVEN AUTOPHONIC LEVELS OF THE PHONEME /a/, RENDERED BY ONE SPEAKER, AND COMPARABLE DATA REPORTED IN TWO OTHER STUDIES

Autophonic Level (db relative)	Frequency (cps)				
	F_0	F_1	F_2	F_3	F_4
0	220	700	1100	1900	2500
-5	160	700	1100	2100	2500
-10	140	500	900	1900	2500
-15	130	600	1000	2000	2700
-20	120	600	1000	2100	2600
-25	120	500	1000	2000	3000
-30	100	500	1000	2000	2900
Lehiste and Peterson (1961) Means of five speakers		665	1145	2520	
Peterson (1961) Means of four speakers	120	760	1065	2550	3570

TABLE II

ESTIMATES OF VOWEL LOUDNESS AT SEVEN SOUND PRESSURE LEVELS (SPL) OR AT SEVEN AUTOPHONIC LEVELS (AL), AND THE SLOPE OF THE LOUDNESS FUNCTION.

Each Entry is the Median of 40 Estimates, Four by Each of 10 S's.

A	B	C							D
No.	Stimulus Series Description	Sound Pressure or Autophonic levels (db)							Slope
1	AL and SPL covaried	50	55	60	65	70	75	80	.4
		5	6.5	8	10	12	15	20	
2	AL constant (-5db),SPL varied	50	55	60	65	70	75	80	.4
		5	7.5	9	10	13	18	20	
3	AL constant (-15db),SPL varied	50	55	60	65	70	75	80	.4
		5	6	8	10	13	20	20	
4	AL constant (-25db),SPL varied	50	55	60	65	70	75	80	.4
		5	8	10	10	12	15	20	
5	Synthesized /a/ Pulse spectrum SPL varied	50	55	60	65	70	75	80	.4
		5	5	8	10	12	15	19	
6	Synthesized /a/ Noise spectrum SPL varied	50	55	60	65	70	75	80	.5
		4	5	8	10	12	15	20	
7	Synthesized formant SPL constant Pitch varied	-30	-25	-20	-15	-10	-5	0	.1
		10	10	10	10	12	13	12.5	
8	SPL constant, AL varied	-30	-25	-20	-15	-10	-5	0	.1
		10	12	10.5	11	13	15	14	
9	AL varied in 5db steps SPL varied in 2db steps	-30	-25	-20	-15	-10	-5	0	.3
		8	10	10	10	15	18	20	
10	1,000 cps tone SPL varied	50	55	60	65	70	75	80	.5
		3	5	5	7	10	15	20	

Figure Caption

Fig. 1. Sound pressure levels produced by the PDR-10 earphone in a 6 cc. coupler for one volt input. (Calibrations with General Radio oscillator 1304-B, graphic level recorder 1521-A, and a Western Electric 640A condenser microphone).

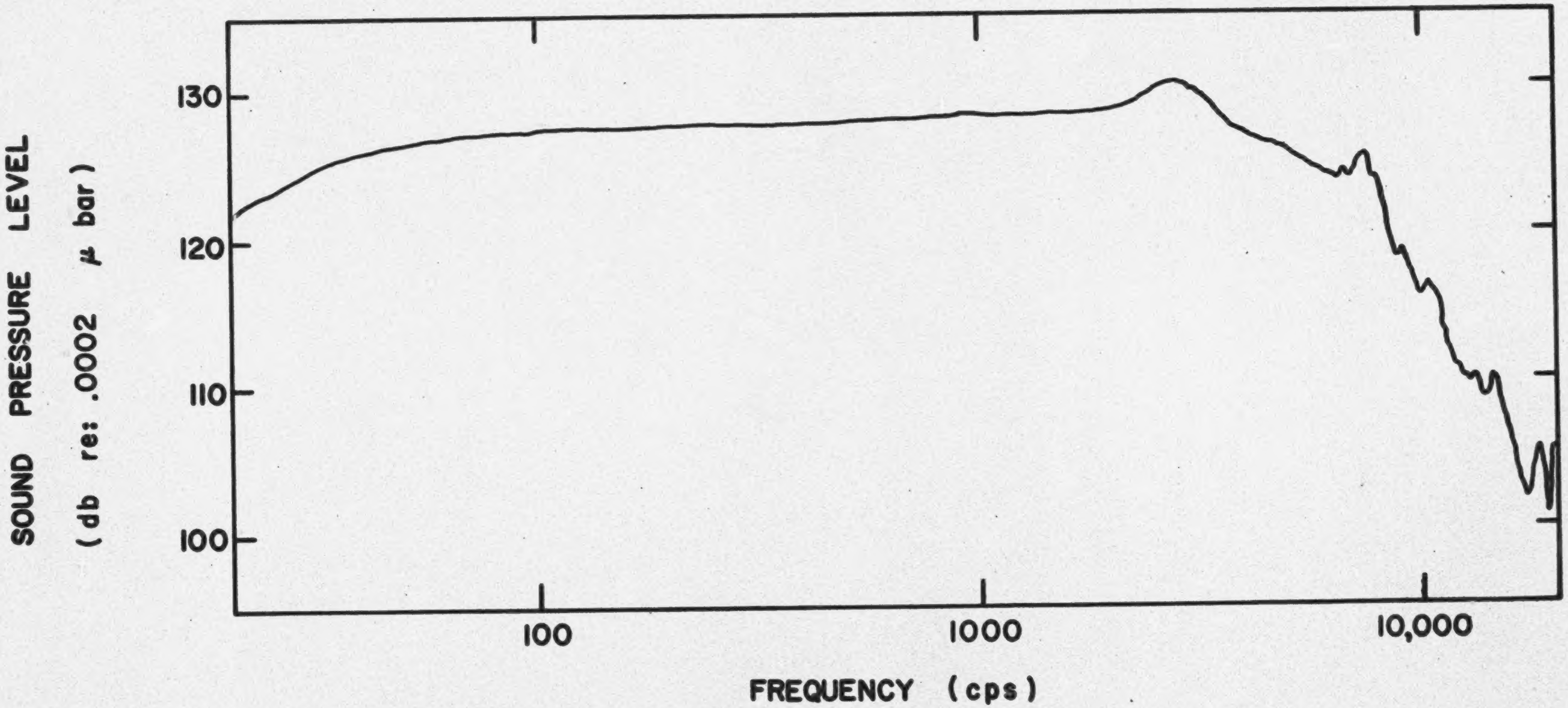
SOUND PRESSURE LEVEL
(db re: .0002 μ bar)
130
120
110
100
100
1000
10,000
FREQUENCY (cps)

Fig. 1

THE EFFECTS OF CHANGING VOWEL PARAMETERS ON PERCEIVED LOUDNESS AND STRESS.

II: SOUND PRESSURE, SPECTRAL STRUCTURE, AND AUTOPHONIC LEVEL

Harlan Lane
The University of Michigan

THE EFFECTS OF CHANGING VOWEL PARAMETERS ON PERCEIVED LOUDNESS AND STRESS. II: SOUND PRESSURE, SPECTRAL STRUCTURE, AND AUTOPHONIC LEVEL

Harlan Lane
Communication Sciences Laboratory
The University of Michigan

The first study in this series (Lane, 19) showed that vowel loudness grow as the 0.4 power of the sound pressure and that this exponent is unaffected by changes in the autophonic level of the stimuli presented. However, when sound pressure was held constant and autophonic level varied there was a slight tendency for numerical estimates to increase as autophonic level increased.

Further clarification of the role of autophonic level in determining vowel loudness awaits the answer to at least these four questions. (1) With sound pressure held constant, changes in autophonic level were shown to affect estimates of vowel loudness. Is this finding merely the result of procedural constraints? The need for a method was indicated which would yield loudness judgments of vowels, varying in autophonic level but constant in sound pressure, without constraining the subject to vary his numerical estimates.

(2) The hypothesis that speech is perceived by reference to articulation was examined in the context of vowel loudness judgments and it was shown that the growth of vowel loudness is not perceived by reference to autophonic level. An alternate interpretation of the mediation hypothesis may be of-

fered, however. Allowing that the slope of the vowel loudness function is not affected by autophonic level, is the intercept thus affected? The need for a method was indicated which would yield loudness judgments of vowels of various autophonic and sound pressure levels, relative to a common standard.

(3) If autophonic level can be shown to influence vowel loudness judgments, how does the magnitude of this effect compare with that for sound pressure level? Is there a large interaction as well? The need for a method was indicated which would permit an assessment of the relative magnitude of effect of several variables operating in concert.

(4) The first experiment showed that, by changing only the fundamental frequency of a single band of harmonics, an increase in loudness judgments could be obtained which was comparable to that effected by changes in autophonic level with sound pressure held constant. This raises the question of whether an effect of autophonic level, if demonstrated, is related to vocal behavior or to some other variable (e.g., changes in sensitivity of the ear as pitch increases.) Does the effect of autophonic level on vowel loudness decrease as the "speech likeness" of the signal is decreased, that is, as the spectral structure of the vowel is degraded? Only under this circumstance may we consider an effect of autophonic level to be peculiar to speech perception.

The present study employs an innovation in the method of magnitude estimation which yields findings that answer these questions. The vowel parameters: autophonic level, intensity, and spectral structure are incorporated in a three dimensional experimental design. The levels of the autophonic variable are 0, 10, 20, and 30 db (relative); the levels of the intensity variable are 50, 60, 70, 80 db (SPL); the levels of the spectral structure variable are fundamental frequency, first formant, first and second formants, and total vowel.

Using the method of magnitude estimation, subjects give loudness estimates, relative to a single standard, for the 64 stimuli defined by this three-space. The effect of the four autophonic levels on the vowel loudness function is then examined, constituting a replication of the first experiment in this series. The effect of autophonic level on loudness estimates with sound pressure held constant at one of four levels is also examined. The greatest power of the method is that, furthermore, it permits an analysis of variance of the loudness estimates, and a comparison of the size of the mean squares for the three main effects and their interactions.

The multidimensional scaling procedure described eliminates many of the constraints entailed by the stimulus series approach to magnitude estimation; it may, however, introduce other response biases. Since the procedure involves numerical estimation of stimuli changing from moment to moment in several dimensions, a difficult task at best (see Stevens and Poulton, 1956), an indication of its validity would seem to be in order prior to exploring the complex problem of vowel loudness. Experiment 1 of the present study employs this technique to determine the loudness function for a 1,000 cps tone. By using relatively simple and well-studied acoustic parameters (duration, rise-time, and intensity level) Experiment 1 permits a comparison of findings obtained with this technique to those obtained with more traditional scaling methods. With this criterion, the validity of the method is established in Experiment 1 and the method is then employed in Experiment 2 for the study of vowel loudness.

Experiment 1

Method

A three-way experimental design was employed. The variables and their levels were: rise time (10, 100 msec.), duration (400, 800, 1600 msec.), and intensity level (50, 60, 70, 80 db, SPL) of a 1,000 cps tone.

To obtain the 24 stimuli, the output of an oscillator (Hewlett-Packard 207A), generating a sine wave at 1,000 cps (calibrated with a Hewlett-Packard Frequency Counter 522B) was sent to an electronic switch (Grason-Stadler 8295119) controlled by an interval timer (Grason-Stadler 471). Time intervals were calibrated with a frequency counter (supra). The nominal rise-time settings on the electronic switch were calibrated by causing the switch to multivibrate and displaying the gated signal on an oscilloscope (Tektronics 533)whose horizontal sweep was synchronized with signal onset. The output of the electronic switch was varied in 10 db steps with a calibrated attenuator (Hewlett-Packard 350A) and recorded at one of four VU levels over a 30 db range (Ampex 300 tape recorder, 7.5 ips). The 24 stimuli appeared in random order at four-second intervals within each of four series. The loudness standard or "modulus" appeared at the beginning of each series (1,000 cps tone, four seconds, 0 VU.)

Each of ten unpracticed undergraduates served as subjects in experimental sessions lasting about 15 minutes. The subject was seated in front of a microphone in an anechoic chamber. He wore a pair of PDR-10 earphones with sponge-Neoprene cushions (MX-41/AR). The experimenter and tape recorder were located in an adjacent room. The playback system was adjusted so that a 1,000 cps tone recorded at 0 VU on the tape recorder produced a signal of 80 db (SPL) at the headphones. Measurements of voltage levels at the receiver terminals were made during tests and later converted to sound pressure levels by means of a receiver calibration. These measurements were made with sustained tones so the results are not dependent upon the time characteristics of the measuring apparatus.

The following instructions were read to the subject.

> "This is an experiment to see how you perceive the loudness of some sounds. I will play four tape recordings to you, each consisting of 24 stimuli. Your task will be to give a numerical estimate of the loudness of each stimulus as it comes along.
>
> "The first sound on each tape is the standard loudness which you are to call "100." Then, after you hear the second sound tell me the numerical value of its loudness, and so on through the tape. Always assign numbers to the stimuli in the same proportion to "100" as their loudness is to the standard. For example if the second stimulus is twice as loud as the standard call it "200," if the second stimulus is half as loud as the standard call it "50." Are there any questions?

"Let's begin. Remember, the first stimulus on each tape is the standard and you are to assign the number "100" to it."

Results and Discussion

Table I shows the median loudness estimates assigned to each of the 24 stimuli. Table II shows the slope of the loudness function. The findings of the present study, employing multidimensional scaling, may be compared to those obtained using the method of magnitude estimation with a stimulus series varying along only one dimension (intensity). A study by Stevens and Poulton (1956) is comparable to the present experiment in all other respects. These atuhors presented 1,000 cps tones varying in intensity from 60 to 75 db (cf. 50-80 db) to eleven unpracticed observers for numerical estimates of loudness. The modulus was 6 db greater than the most intense stimulus (cf. 0 db) and was assigned the value of 100 by the experimenter. The duration and rise-time of the stimuli were approximately 1600 and 10 msec., respectively. As in the present study, Stevens and Poulton found that a straight line provides a good fit to the logarithm of the median estimates as a function of sound pressure level; the slope of this line is approximately 0.43. This exponent for the loudness function is identical with that obtained in the present study at 10 msec. rise-time, 1600 msec. duration—this, despite the critical difference in procedure.

The method used in the present study also permits an assessment of the effects of rise-time and duration on loudness. Table I shows that rise-time has no consistent effect on loudness judgment under the conditions of this experiment. However, there is an obvious increase in loudness estimates as duration is increased. Figure 1 presents equal loudness contours as a function of duration for the results of the present study and two other investigations. The numerical estimates shown in Table I were averaged over rise-times and converted to decibels relative to the loudness of the stimulus with duration 400 msec., intensity 50 db (SPL). These decibels of loudness were then converted to equivalent db of sound pressure level by the formula: $L_{db} = 0.5\ SPL_{db}$ and plotted in Fig. 1. The parameter of the contours are the four levels of the intensity variable. In this form the data may be compared to those obtained by Munson (1947), using an entirely different procedure. Munson presented each of five observers with a sequence of two 1,000 cps tones. The first tone, with a loudness level of 70 phons, was of variable duration. The second tone had a duration of one second but its intensity was varied. The subject reported which tone was the louder. The equal-loudness contour shown in Fig. 1 gives the intensity level of the one-second tone at which it was called louder 50% of the time when paired with another tone of indicated duration. Finally, Fig. 1 shows the effect on the threshold of audibility of changes in the duration of pure tones (Miskolczy-Fodor, 1959). Each point is the mean threshold shift obtained with 40 normal ears at three frequencies.

In view of the similarities among the findings obtained in the present experiment and those reported by other authors, it was concluded that the method employed had sufficient validity to warrant its use in an investigation of the parameters of vowel loudness.

Experiment 2

Method

Four autophonic levels of the phoneme /a/ were generated at 10 db intervals by one speaker.[2] A 500 msec. sample of each level was obtained with an electronic switch (Grason-Stadler 8295119; rise-time, 100 msec.), controlled by a calibrated interval timer (Grason-Stadler 471). The gated signals were recorded on magnetic tape (Ampex 300 tape recorder, 7.5 ips) and spectrograms were prepared (Western Electric BTL-2 spectrograph). The acoustic parameters of these vowels, which constituted the autophonic variable of the present study, are given in Table III. The fundamental frequency of each vowel did not vary by more than 2 cps, nor the average speech power by more than 0.5 db, during the 500 msec. sample.

Amplitude sections of each vowel were made with the sound spectrograph and the fundamental and formant frequencies identified; we used as a guide the data reported by Peterson (1961), and by Lehiste and Peterson (1961). In order to obtain the four levels of the spectral structure variable: fundamental frequency only (F_0), first formant only (F_1), first and second formants (F_1 and F_2) and total vowel (T), each of the four autophonic levels was band-pass filtered as appropriate. Two filters (Krohn-Hite 310-AB) were connected in series and interposed between a playback tape recorder and the sound spectrograph. The experimenter made successive adjustments of the filter settings until an amplitude section of the filtered signal approximated the corresponding segment of the amplitude section for the total vowel. The filter settings employed for the "total" vowel were 90-4,000 cps. (The low cutoff was slightly lower than the lowest fundamental frequency and the high cutoff was 500 cps greater than the highest frequency displayed on the spectrograph.)

In order to obtain the four levels of the sound pressure variable, the average speech power of each of the 16 signals was displayed on a recording oscillograph (Minneapolis Honeywell Visicorder) which was calibrated so that the peak speech power could be read within 0.5 db. Using this record as a guide, various amounts of attenuation were introduced at the output of the playback tape recorder and the 16 signals were recorded on a second recorder at 0, 10, 20, and 30 db below zero VU.

The 64 stimuli obtained in this manner were presented in irregular order in four successive series to each of 15 unpracticed observers. Each series began with a five-second 1,000 cps tone recorded at 0 VU, which served as the standard. The subject was seated in front of a microphone in an anechoic chamber; he wore a pair of PDR-10 earphones mounted in MX-41/AR cushions. (The frequency response of the earphones has been presented previously [Lane, 19]).

The experimenter was located in an adjacent control room. The output of the tape recorder was sent to a transistorized earphone amplifier with high signal to noise ratio and flat frequency response over the range 20-20,000 cps, and then to the subject's headset. The playback system was adjusted so that a 1,000 cps tone recorded at 0 VU would produce a sound pressure level of 80 db at the earphones. The instructions to the subject were identical to those employed in Experiment 1 (supra).

Results and Discussion

An analysis of variance of the vowel loudness estimates is presented in Table IV to permit comparison of the mean squares for the sound pressure, spectral structure, and autophonic variables. Clearly, the major determinant of vowel loudness estimates is the sound pressure level of the vowel. Figure 2 shows the vowel loudness function at each of the four levels of spectral structure, the variable with the second largest effect. It will be seen that degrading the spectral structure of the vowel has only a very slight effect on the slope of the loudness function, although it has a large effect on the intercept of this function. Concerning the slope of these functions, it is noteworthy that vowel loudness grows at approximately the same rate as pure tone loudness as a function of sound pressure level. The exponent (slope) of 0.4 may at first seem a remarkable departure from the slope of the sone scale for loudness (0.6), but actually the flattening of the function is not unexpected under the conditions of this experiment. Stevens and Poulton (1956) have also obtained slopes of about 0.4 with untrained observers giving numerical estimates relative to a standard at the top of the stimulus series. A complementary procedure for determining the slope of the loudness function, the method of magnitude production (Stevens, 1958), tend to give somewhat higher exponents. The slope of the sone scale for loudness is, therefore, a best estimate. The important point for the present discussion is that the dynamics of vowel loudness does not differ appreciably from that of pure tone loudness.

The effect of spectral structure on the intercept of the loudness function is as surprising as it is dramatic. The effect may be observed more clearly in Fig. 3, which plots median estimates of loudness as a function of spectral structure with sound pressure level as the parameter. If the effect of spectral structure may properly be considered an illusion peculiar to speech signals, it is particularly curious that the total /a/ should be judged louder than the first and second formants combined, since the latter are usually adequate for vowel recognition. The effect of spectral structure on loudness was observed at all levels of the autophonic variable. Acoustic and psycho-acoustic considerations are too complex to permit an accurate correction of these curves for the frequency response of the ear. It is unlikely that the effect can be accounted for entirely by the increasing sensitivity of the ear at higher frequencies. The correction may be assumed to be of the order of only a few db in view of the filter settings employed for the preparation of the levels of spectral structure. Furthermore, equal loudness contours tend to flatten as sound pressure level is increased, whereas the loudness illusion shown in Fig. 3 increases at higher sound pressure levels; this is the sound pressure by spectral structure interaction shown in Table IV.

Table IV shows that autophonic level had only a very small effect on loudness estimates compared to that of sound pressure level and spectral structure. Figure 4 shows the vowel loudness function with autophonic level as a parameter. The slope and intercept of the function are only slightly affected by autophonic level. The loudness functions do diverge at the lowest intensity, which is graphic evidence of the sound pressure by autophonic level interaction shown in Table IV. All of the autophonic level by sound pressure interaction takes place at one level of the spectral structure variable: the total /a/. (This is the second order interaction effect shown in Table IV). The finding that autophonic level affects loudness only so long as the total vowel is judged lends credence to the notion that this is an effect of speech perception and not an artifact. This interaction may be seen more clearly in Fig. 5, where loudness estimates of the total vowel are shown as a function of autophonic level with sound pressure level as the parameter. At the lowest intensity level, the lowest autophonic level is "underestimated," whereas at the highest intensity level, the highest autophonic level is "overestimated."

We may now answer the question, does autophonic level influence vowel loudness? The answer is a conditional yes. Autophonic level does not influence the growth of loudness as a function of sound pressure. Furthermore, autophonic level does not contravene or even attenuate the effects of sound pressure level on loudness. However, when sound pressure and autophonic level are correlated, there appears to be some effect of autophonic level at extreme values. The effect is probably peculiar to speech signals, since it is not observed when the spectral structure of the vowel is degraded.

In terms of the engineering application discussed by Lehiste and Peterson (1959), the value of providing autophonic level information to aid in loudness judgments by an automatic speech recognizer may be seriously questioned. The findings of the present study also question the validity of the mediation hypothesis in the context of vowel loudness judgments (see Lane, 196).

Summary

Several questions were raised concerning the role of autophonic level in determining the slope and intercept of the vowel loudness function. An adaptation of the method of magnitude estimation was conceived which eliminated certain biases of traditional ratio-scaling techniques, and permitted an assessment of the relative magnitude of effect of sound pressure level and autophonic level in determining judgments of vowel loudness.

The validity of the method was tested in an experiment on the loudness of 1,000 cps tones that were simultaneously varied in intensity, duration, and rise-time. Ten subjects gave numerical estimates of the relative loudness of stimuli drawn from this three-dimensional space at random. The findings were comparable to those obtained in other investigations of the loudness function and the effects of duration on loudness.

The multidimensional ratio-scaling technique was therefore employed in a study of the contribution of three variables to vowel loudness judgments: sound pressure level, autophonic level, and spectral structure.

1. Autophonic level does not influence the growth of vowel loudness as a function of sound pressure level.

2. Relative to the effect of sound pressure level the effect of autophonic level on vowel loudness is negligible.

3. When sound pressure level and autophonic level are correlated, the latter has some effect at extreme values.

4. Over the range of sound pressure levels employed, the total vowel is always judged louder than the first or first and second formants when they are at equal sound pressure levels.

Notes

1. The assistance of Mr. D. R. Brinkman is gratefully acknowledged. This research was performed persuant to a contract with the United States Office of Education, Language Development Section.

2. Licklider and Miller (1951) give 40 db as the range of average speech power between the loudest and the weakest vocalizing possible.

References

Lane, H. L. The effects of changing vowel parameters on perceived loudness and stress. I: Does autophonic level affect the loudness function? (Submitted for publication in the J. exp. Psychol.).

Lehiste, I. and Peterson, G.E. Transitions, glides and dipthongs. J. Acoust. Soc. Am., 33, 1961, 268-277.

Licklider, J. C. R. and Miller, G.A. The perception of speech. Ch. 26 in: Stevens, S. S. (ed.) Handbook of Experimental Psychology. John Wiley, New York, 1951.

Miskolczy-Fodor, F. Relation between loudness and duration of tonal pulses. I. Response of normal ears to pure tones longer than click pitch threshold. J. Acoust. Soc. Am., 31, 1959, 1128-1134.

Munson, W. N. The growth of auditory sensation. J. Acoust. Soc. Am., 1947, 584-591.

Peterson, G.E. Parameters of vowel quality. J. Sp. Hear. Res., 4, 1961, 10-29.

Stevens, S. S. Problems and methods of psychophysics. Psychol. Bull., 54, 1958, 177-196.

Stevens, S. S. and Poulton, E. C. The estimation of loudness by unpracticed observers. J. exp. Psychol., 51, 1956, 71-78.

TABLE I

LOUDNESS OF A 1,000 CPS TONE AS A FUNCTION OF DURATION, RISE-TIME AND SOUND PRESSURE LEVEL. EACH CELL ENTRY IS THE MEDIAN OF 40 ESTIMATES, FOUR BY EACH OF TEN UNTRAINED OBSERVERS

Duration (msec.)	Rise Time (msec.)	Sound Pressure Level (db re.0002 μ bar)			
		80	70	60	50
400	10	85	35	25	20
	100	85	40	25	13
800	10	95	50	25	20
	100	85	50	25	15
1600	10	100	58	30	25
	100	90	58	35	25

TABLE II

SLOPES OF THE LOUDNESS FUNCTION OF A 1,000 CPS TONE (LEAST SQUARES FIT TO MEDIAN ESTIMATES) AT THE LEVELS OF THE DURATION AND RISE TIME VARIABLES

Duration (msec.)	Rise Time (msec.)	Slope
400	10	0.4
	100	0.5
800	10	0.5
	100	0.5
1600	10	0.4
	100	0.4

TABLE III

ACOUSTIC PARAMETERS OF FOUR AUTOPHONIC LEVELS OF THE VOWEL PHONEME /a/, AND COMPARABLE DATA FROM TWO OTHER STUDIES

Autophonic level (db relative)	Frequency (cps)				Formant Amplitude (db relative)		
	F_0	F_1	F_2	F_3	F_1	F_2	F_3
0	220	700	1100	1900	0	-2	-17
-10	140	500	900	1900	0	-5	-12
-20	120	600	1000	2100	0	-2	-25
-30	100	500	1000	2000	0	-3	-20
Peterson (1961)	120	760	1065	2550	0	0	-16
Lehiste and Peterson (1961)		665	1145	2520			

TABLE IV

ANALYSIS OF VARIANCE OF ESTIMATES OF VOWEL LOUDNESS

Source	Degrees of Freedom	Mean Squares	F
Sound pressure level (A)	3	628,300	1,603.
Spectral structure (B)	3	217,631	555.
Autophonic level (C)	3	2,580	6.6
A x B	9	20,585	53.
B x C	9	4,625	12.
A x C	9	988	3.5*
A x B x C	27	1,381	3.5
Within cells	3,776	392	

*Significant at the .01 level of confidence. All other F-ratios are significant at the .001 level.

Figure Captions

Fig. 1. Equal loudness contours for duration obtained with three procedures. Magnitude estimation: (this study); each point is the mean of 40 loudness estimates transformed to equivalent decibles of sound pressure. Equal loudness (Munson [1947]). Each point shows the intensity level of a one-second tone at which it was called louder 50 per cent of the time when paired with another tone of indicated duration.

Threshold shift: (Miskolczy-Fodor [1959]); each point is the mean threshold shift (db) obtained with 40 normal ears at three frequencies.

Fig. 2. The effect of spectral structure on the loudness function. Each point is the mean of 240 loudness estimates, 16 by each of 15 untrained observers. (F_0 = fundamental frequency; F_1 = first formant; F_1, F_2 = first and second formants; T = total vowel, filtered 90-4,000 cps.)

Fig. 3. Loudness estimates assigned to various components of the phoneme /a/ presented at four sound pressure levels. Each point is the mean of 240 determinations.

Fig. 4. The effect of autophonic level on the loudness function. Each point is the mean of 240 numerical estimates, 16 by each of 15 untrained observers.

Fig. 5. Loudness estimates assigned to four autophonic levels of the phoneme /a/ presented at four sound pressure levels. Each point is the mean of 240 determinations.

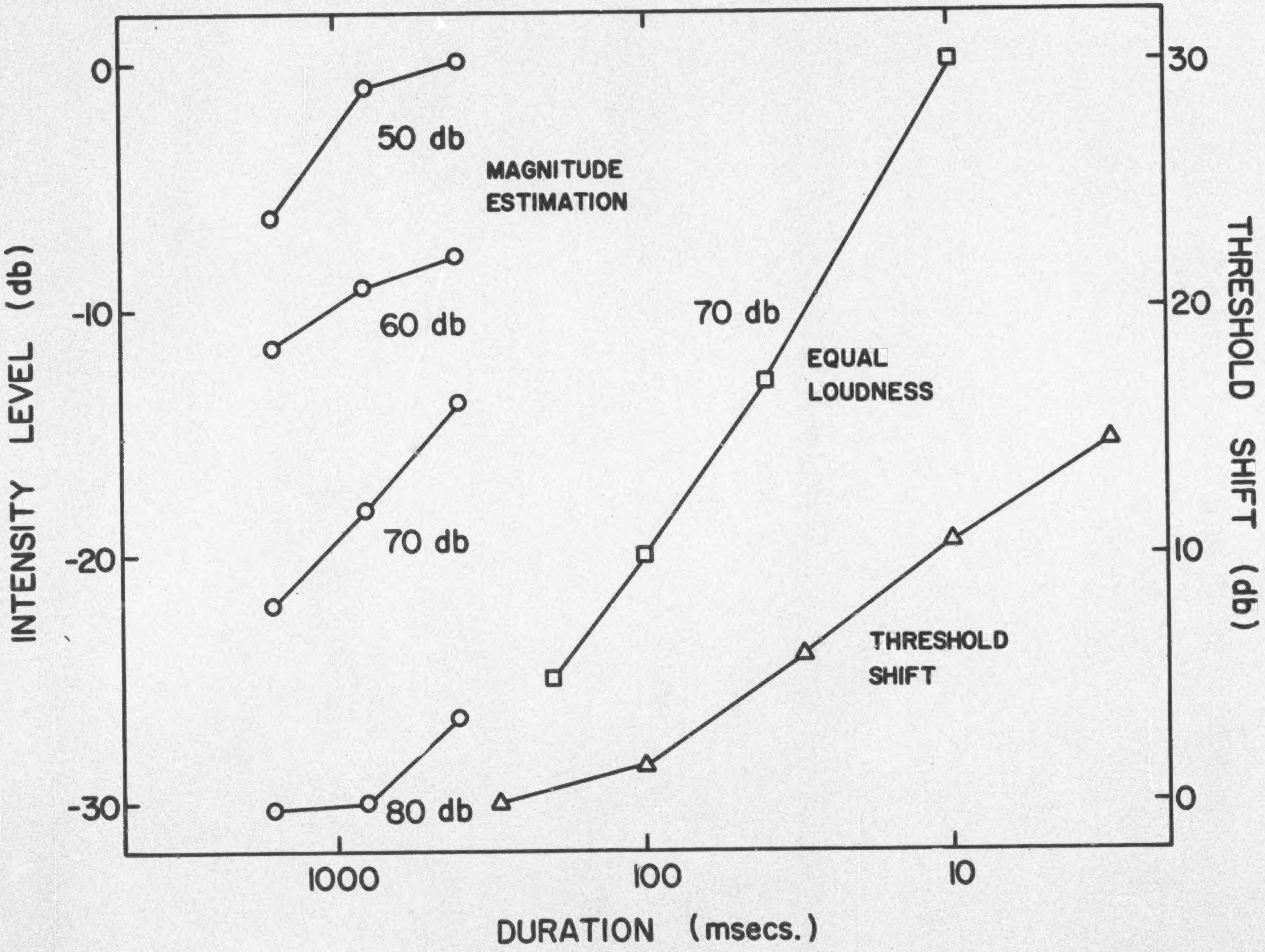
INTENSITY LEVEL (db)
0
-10
-20
-30
THRESHOLD SHIFT (db)
30
20
10
0
1000
100
10
DURATION (msecs.)
50 db
60 db
70 db
80 db
MAGNITUDE ESTIMATION
70 db
EQUAL LOUDNESS
THRESHOLD SHIFT

Fig. 1

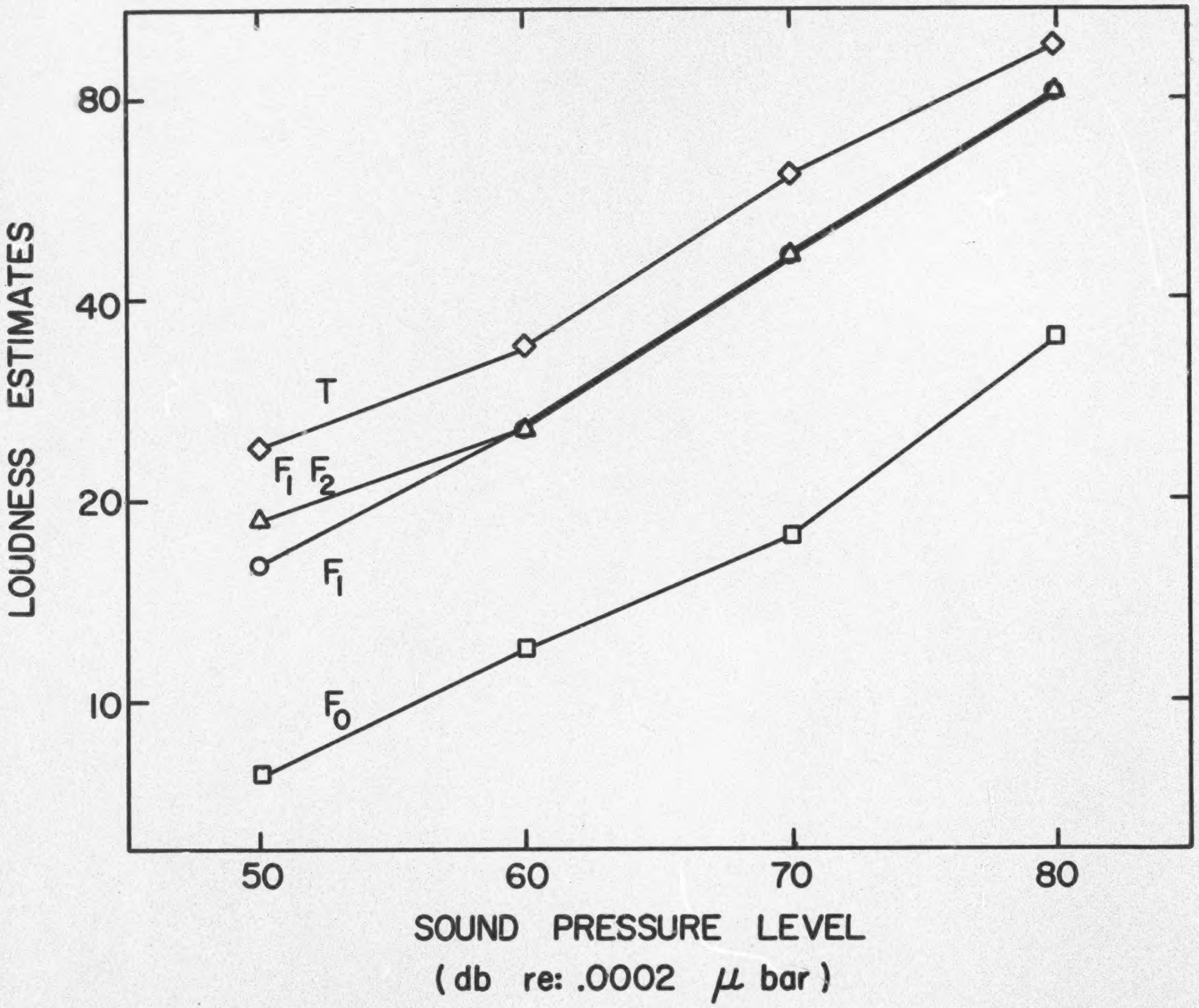
LOUDNESS ESTIMATES
80
40
20
10
50
60
70
80
SOUND PRESSURE LEVEL
(db re: .0002 μ bar)
T
$F_1 F_2$
F_1
F_0

Fig. 2

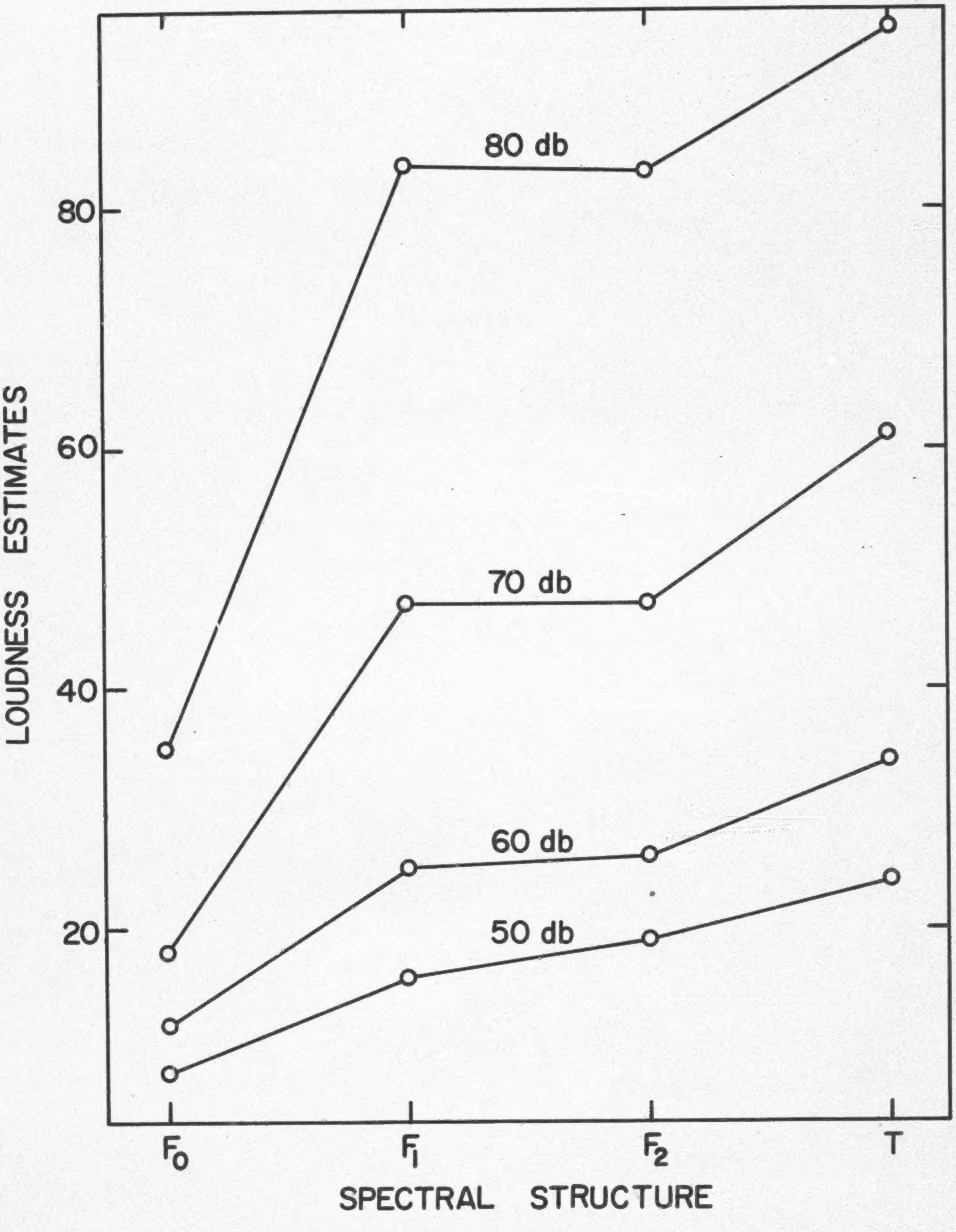
LOUDNESS ESTIMATES
80
60
40
20
80 db
70 db
60 db
50 db
F_0
F_1
F_2
T
SPECTRAL STRUCTURE

Fig. 3

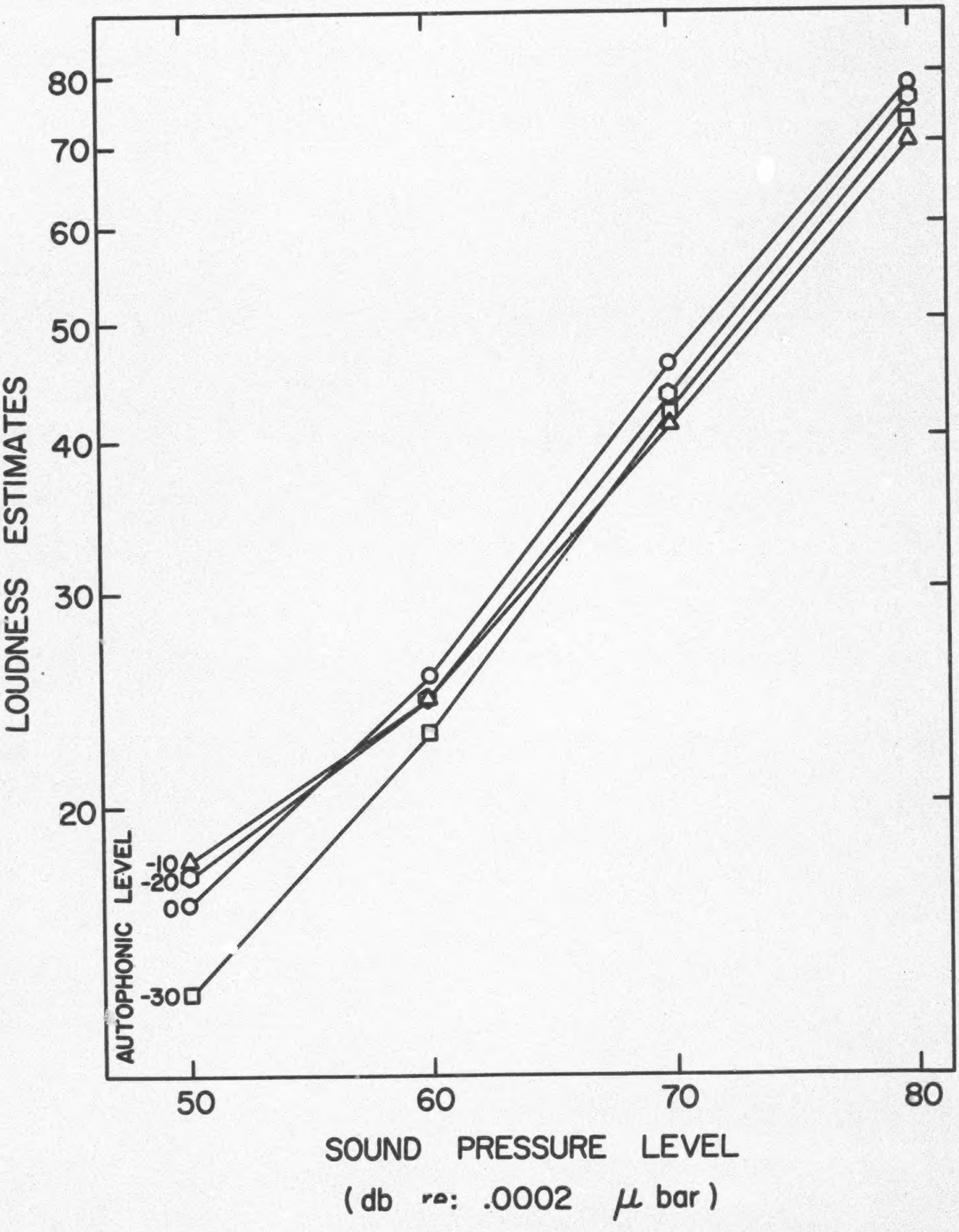
LOUDNESS ESTIMATES
80
70
60
50
40
30
20
AUTOPHONIC LEVEL
-10
-20
0
-30
50
60
70
80
SOUND PRESSURE LEVEL
(db re: .0002 μ bar)

Fig. 4

120

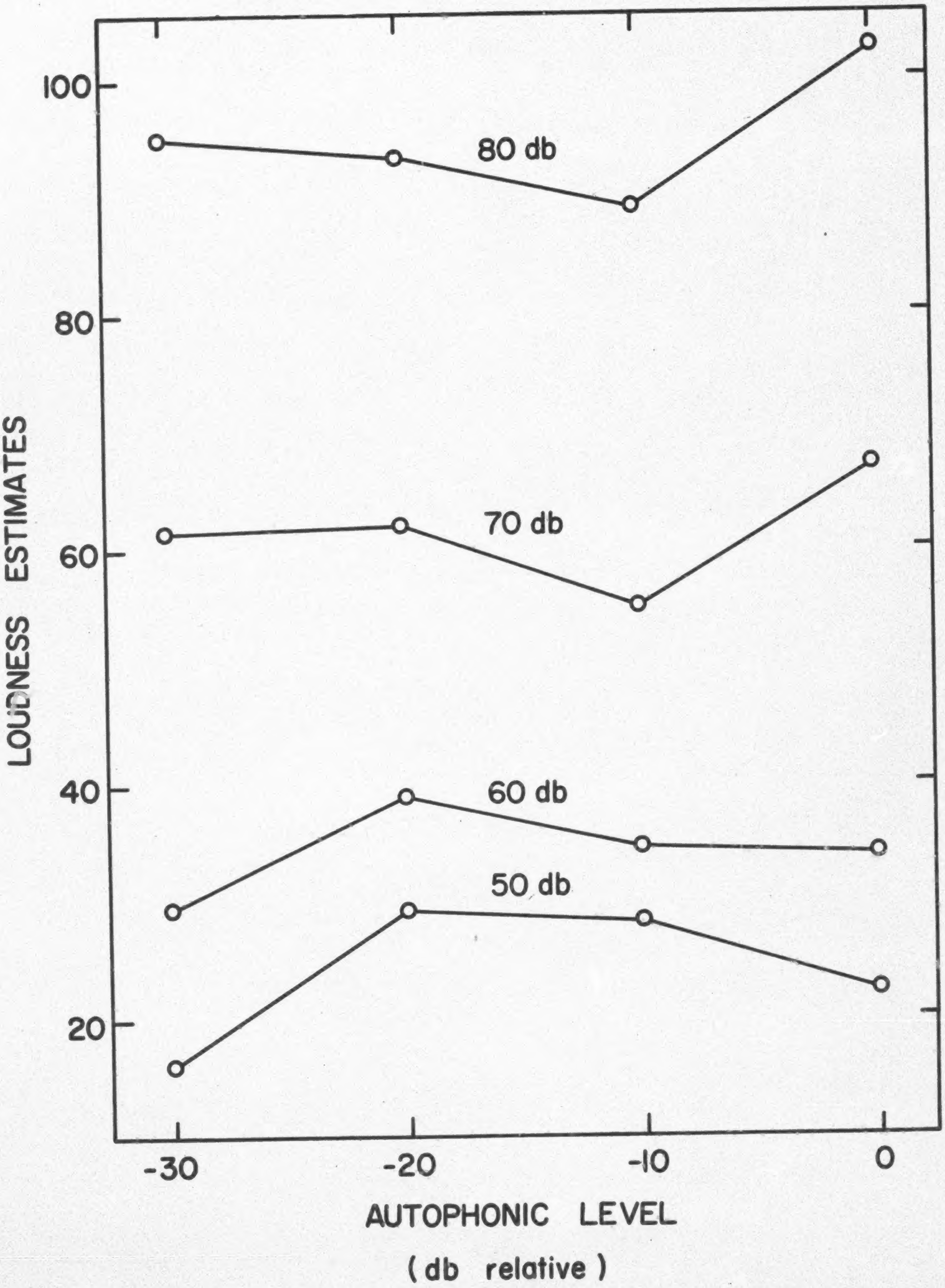
LOUDNESS ESTIMATES
100
80
60
40
20
80 db
70 db
60 db
50 db
-30
-20
-10
0
AUTOPHONIC LEVEL
(db relative)

Fig. 5

THE EFFECTS OF CHANGING VOWEL PARAMETERS ON PERCEIVED LOUDNESS AND STRESS.

III: VOCAL MATCHING OF STRESS PATTERNS

Harlan Lane
The University of Michigan

THE EFFECTS OF CHANGING VOWEL PARAMETERS ON PERCEIVED LOUDNESS AND STRESS
III: VOCAL MATCHING OF STRESS PATTERNS

Harlan Lane
Communication Sciences Laboratory
The University of Michigan[1]

The hypothesis that speech is perceived by reference to articulation was examined in the context of vowel loudness judgments by the first two studies in this series. It was shown that the dynamics of sensory magnitude in speech production (the autophonic scale) differ greatly from those in speech reception (the vowel loudness scale). Contrary to the opinion of other authors,[2] the subject does not seem to perceive the loudness of a vowel in terms of the autophonic level which was required to produce it. Indeed, autophonic level was shown to have a negligible effect on the slope and intercept of the vowel loudness function.

The present study deals a coup de grace to the contention that vowel loudness is perceived in terms of vocal effort and serves, as well, to verify the subjective scales for these two processes. The subject is given the task of imitating some iambic and trochaic stress patterns. If these patterns are perceived in terms of the vocal effort required to produce them, then changes in stress should be linearly related to changes in the vocal matching response. If, on the other hand, the reception and production of stress have different operating characteristics, then the outcome of cross-modality matches should reflect the dynamics of sensory magnitude in the two modalities and reveal these to be related nonlinearly.

In general, if two continua are governed by the equations

$$\psi_1 = \Phi_1^m \quad \text{and} \quad \psi_2 = \Phi_2^n$$

and if the psychological values ψ_1 and ψ_2 are equated at various levels, it follows that the stimulus values Φ_1 and Φ_2 should stand in the relation

$$\log \Phi_1 = (n/m) \log \Phi_2.$$

Thus, cross-modality matches yield a function that is a straight line when plotted in log-log coordinates and has a slope given by the ratio of the exponents of the power laws governing the two modalities (cf. Stevens, 1960). In particular, Lane, Catania, and Stevens (1961) have shown that the power law governing the speaker's perception of his own vocal level, the autophonic response, has an exponent of 1.1. Lane (196) has shown that vowel loudness grows as the 0.4 power of sound pressure level. Therefore, we may predict that the decibel differences in intensity within the stress patterns will be related to decibel differences in the matching response by a straight line with slope 0.4/1.1 = 0.34. This prediction is based on the assumption that the subject does not redintegrate to speaking when he is listening. The contrary assumption leads, as indicated, to the prediction that the matching function will have a slope of 1.0.

Method

The word /ba/, rendered by the experimenter, was recorded on a loop of magnetic tape with one channel of a four-channel tape recorder (Ampex 300-4). A narrow band spectrogram (prepared on a Western Electric BTL-2 spectrograph) showed that the fundamental frequency of 120 cps did not vary by more than 2 cps during the vowel "steady state" and that the duration of the entire signal was 340 msec. The signal was also processed by an average speech power circuit (integrating time, 10 msec.) and the output recorded on a calibrated oscillograph (Minneapolis-Honeywell Visicorder). This trace of the average speech power as a function of time showed that the amplitude of the vowel did not vary by more than 0.5 db during the steady state. By connecting the output of channel 1 of the tape recorder to the input of channel 2 and the output of channel 2 to the input of channel 3, the original signal was copied on the third channel after a 100 msec.delay. The recorded signals on channels 1 and 3 were mixed electronically during playback and recorded on channel 4, giving the stimulus for this experiment: /ba-ba/. The signals on the magnetic tape loop were then played back repeatedly and the stimulus series prepared. The original signal on channel 1 served to trigger two electronic timers (Grason Stadler 471) and an electronic switch (Grason Stadler 829S119), which were controlled in series. The first timer introduced a delay of 100 msec., after which it triggered the second timer. This timer closed the electronic switch, sending the first /ba/ from the output of channel 4 to an attenuator (Hewlett-Packard 350A). After 350 msec., the switch returned to the normally closed position with the effect that the second /ba/ from channel 4 was sent to a second attenuator. The two "syllables" were sorted in this manner to permit adjustment of their relative intensity with the calibrated attenuators. The settings of the attenuators were adjusted according to a protocol during the five-second inter-stimulus interval determined by the magnetic tape loop. The outputs of the attenuators were sent through an electronic mixer to a second continuous tape recording (Ampex 350).

Two stimulus series were prepared in this manner. In the first series (A) the first stimulus in each pair had a constant VU level (-15 db) while the second was recorded three times in irregular order at each of the following levels: -30, -25, -20, -15, -10, -5, 0 db below 0 VU. In the second series (B), the second stimulus was held constant and the first stimulus varied as in (A). The procedure yielded 42 disyllables, 21 with iambic and 21 trochaic stress.

Seven male and two female undergraduates served in sessions lasting approximately 30 minutes. The subject was seated in an anechoic chamber in front of a microphone (Altec 633A); the experimenter was located in an adjacent control room, along with a tape recorder that served to present the stimulus series and to record the subjects' matching responses. The stimulus series were presented over a loudspeaker at a comfortable listening level, but the exact range of intensities was not determined. The following instructions were read to the subject:

> "You will hear pairs of sounds that constitute a stress pattern. During the pause following each pair of sounds attempt to immitate what you heard. Do not move your head during the course of the experiment."

The tape recordings of the subjects' matching responses were processed with an average speech power circuit and the output recorded on an oscillograph. The decibel difference between the two responses was determined to an accuracy of ±.25 db by measuring the distance (in mm) between the peak average speech power of the two responses and converting to decibels with a recorder calibration. Decibel differences among the stimulus pairs were also verified by this procedure. The recordings for two subjects were processed to yield duration and pitch data, as well. Durations were measured from the width of the oscillograph tracing of the average speech power. An adjacent channel of the oscillograph was controlled by the output of a pitch meter (SRL) which applied a d-c voltage that was proportional to the fundamental frequency of the subject's vocal response.

Results and Discussion

Figure 1 shows the function that relates the intensity ratio of the stress pattern to that of the matching response. Two functions are shown: one for intensity changes below the standard and one for those above the standard. The slope of the straight line of best fit (method of least squares) is indicated in each case. Based on the assumption that stress matching would reflect the dynamics of vowel loudness and autophonic level, it was predicted that the matching function would be well-represented by a straight line in a log-log plot with a slope of approximately 0.34. The data shown in Fig. 1 validate this prediction. Clearly, the listener does not perceive the speech stimuli in the same way as he perceives his own vocal behavior. If he did, the matching function would have a slope of 1.0 instead of the obtained slope of 0.34. The obtained slope reflects the dynamics of two different receptor processes: speech and hearing. It appears that S. Jones was in error when he wrote:

"The listener refers what he hears to how he would say it. Thus he translates exteroceptor into proprioceptor sensations, the kinaesthetic memory serving as stimulus" (Jones, 1932).

It must be allowed, however, that the isolated and repetitive sounds of the present experiment may not constitute the type of speech sample which Jones envisioned in his statement. The stimuli in the present study were intentionally held constant in pitch and duration, although changes in these parameters are normally correlated with changes in linguistic stress in English (Fry, 1955, 1958). Figure 2 shows that, despite the fact that the members of each disyllable were of the same pitch and duration, the subjects tended to match higher pitches and longer durations (as well as greater intensities) to the stimuli with greater stress. It is not clear to what extent this finding is due to the mechanics of speech and to what extent it is due to the perception of linguistic stress.

It is interesting to consider the implications of the psychophysical data shown in Fig. 1 for the problem of second-language learning. If a student is required to render a "natural" production of a non-native stress pattern, his initial attempts may be expected to be quite inaccurate. In the particular case of a disyllable with, for example, an intensity ratio of 10 db, the echoic response will have a ratio of only 3.5 db. From the point of view of an observer, a 4 db change in stimulus loudness is matched by the student with a 1.4 db change in response loudness. At least one language laboratory has observed that its students reliably "underestimate" the required loudness ratio when learning to render foreign stress patterns.[3] The reason is now clear.

Summary

The method of cross-modality matching was employed to reveal some of the factors operating in the perception of linguistic stress and to verify the form and exponent of the subjective scales for vowel loudness and autophonic level. Nine subjects gave vocal matches of 42 iambic and trochaic stress patterns of the disyllable /ba-ba/. The function relating intensity ratios of the stress patterns to those of the matching response had the form and exponent which was predicted on the assumption that stress matching is governed by the dynamics of vowel loudness and autophonic level. The data do not support the hypothesis that linguistic stress is perceived by reference to articulation.

References

Draper, M. H., Ladefoged, P. and Whitteridge, D. Respiratory muscles in speech. J. Sp. Hear. Res., 2, 1952, 16-27.

Fry, D. B. Experiments in the perception of stress. Lang. and Sp., 1, 1958, 126-152.

Fry, D. B. Duration and intensity as physical correlates of linguistic stress. J. Acoust. Soc. Am., 27, 1955, 765-768.

Jones, S. The accent in French—what is accent? Le Maitre Phonetique, 40, 1932, 74-75.

Ladefoged, P. Syllables and stress. Miscellanea Phonetica, 3, 1958, 1-14.

Lane, H. L. The effects of changing vowel parameters on perceived loudness and stress. I: Does autophonic level affect the loudness function? (submitted for publication in the J. exp. Psychol.)

Lane, H. L. The effects of changing vowel parameters on perceived loudness and stress. II: sound pressure, spectral structure, and autophonic level. (submitted for publication in the J. exp. Psychol.)

Lane, H. L., Catania, A. C., and Stevens, S. S. Voice level: autophonic scale, perceived loudness, and effects of sidetone. J. Acoust. Soc. Am., 33, 1961, 160-167.

Lehiste, I. and Peterson, G. E. Transitions, glides, and dipthongs. J. Acoust. Soc. Am., 33, 1961, 268-277.

Stevens, S. S. The psychophysics of sensory function. Amer. Scient., 48, 1960, 226-253.

Footnotes

1. This research was performed pursuant to a contract with the United States Office of Education, Language Development Section. The assistance of Mr. Giles Peterson is gratefully acknowledged.

2. Draper, Ladefoged, and Whitteridge (1952); Ladefoged (1958); Lehiste and Peterson (1959).

3. Communicated by F. R. Morton, Director, University of Michigan Language Laboratory.

FIGURE CAPTIONS

Fig. 1. Intensity ratio of the matching response as a function of the intensity ratio of the criterion stress pattern. The decibel difference associated with the matching response is given by the left hand curve for intensity changes below the standard and by the right hand curve for intensity changes above the standard. Straight lines have been fit visually to the means of 54 determinations, six by each of nine observers. The slopes shown were obtained by the method of least squares.

Fig. 2. Intensity, duration, and frequency ratios for the matching responses of two subjects as a function of the intensity ratio of the criterion stress pattern. Each point is the mean of 12 determinations.

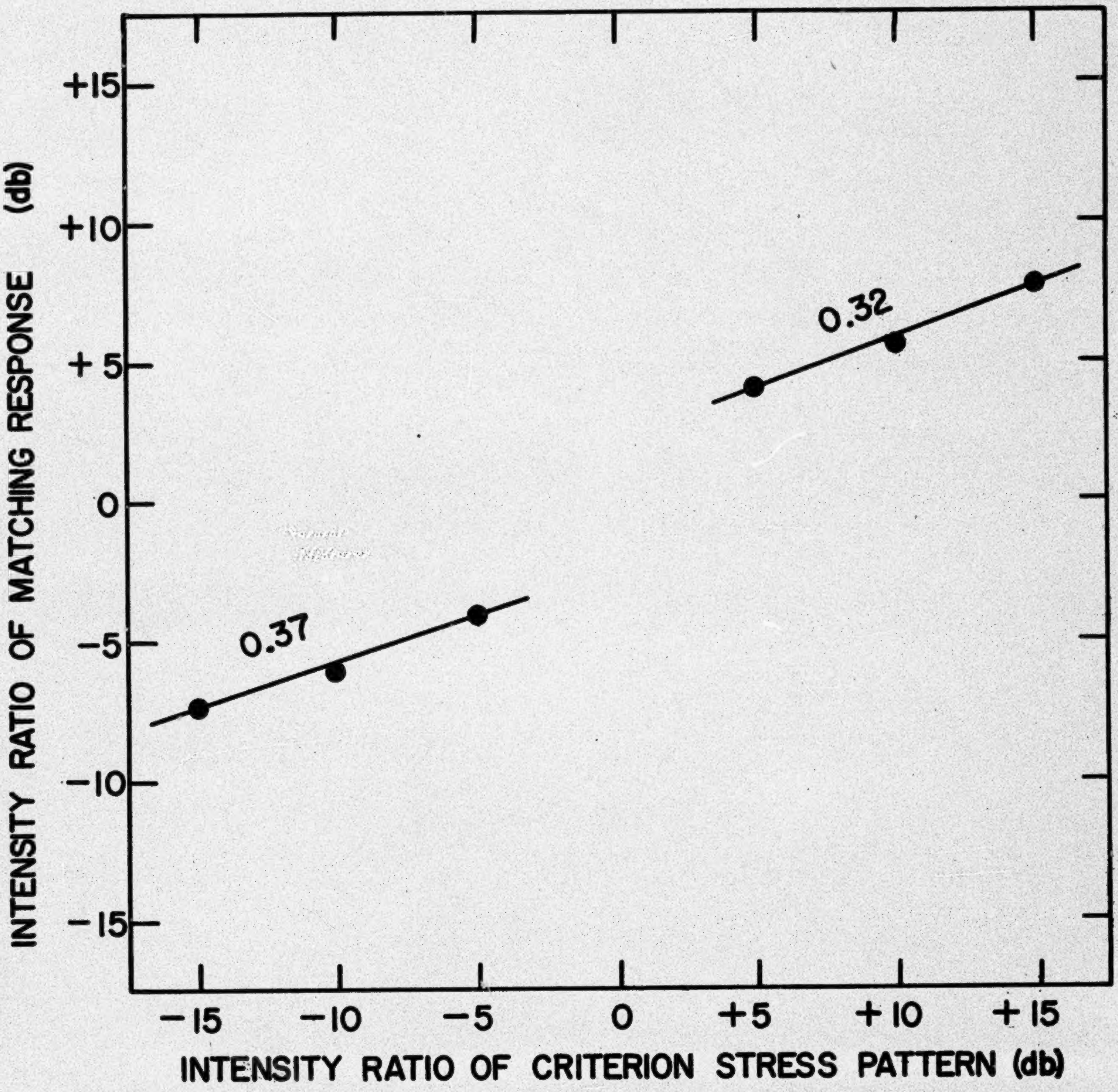
INTENSITY RATIO OF MATCHING RESPONSE (db)
+15
+10
+ 5
0
-5
-10
-15
0.32
0.37
−15
−10
−5
0
+5
+10
+15
INTENSITY RATIO OF CRITERION STRESS PATTERN (db)

Fig. 1

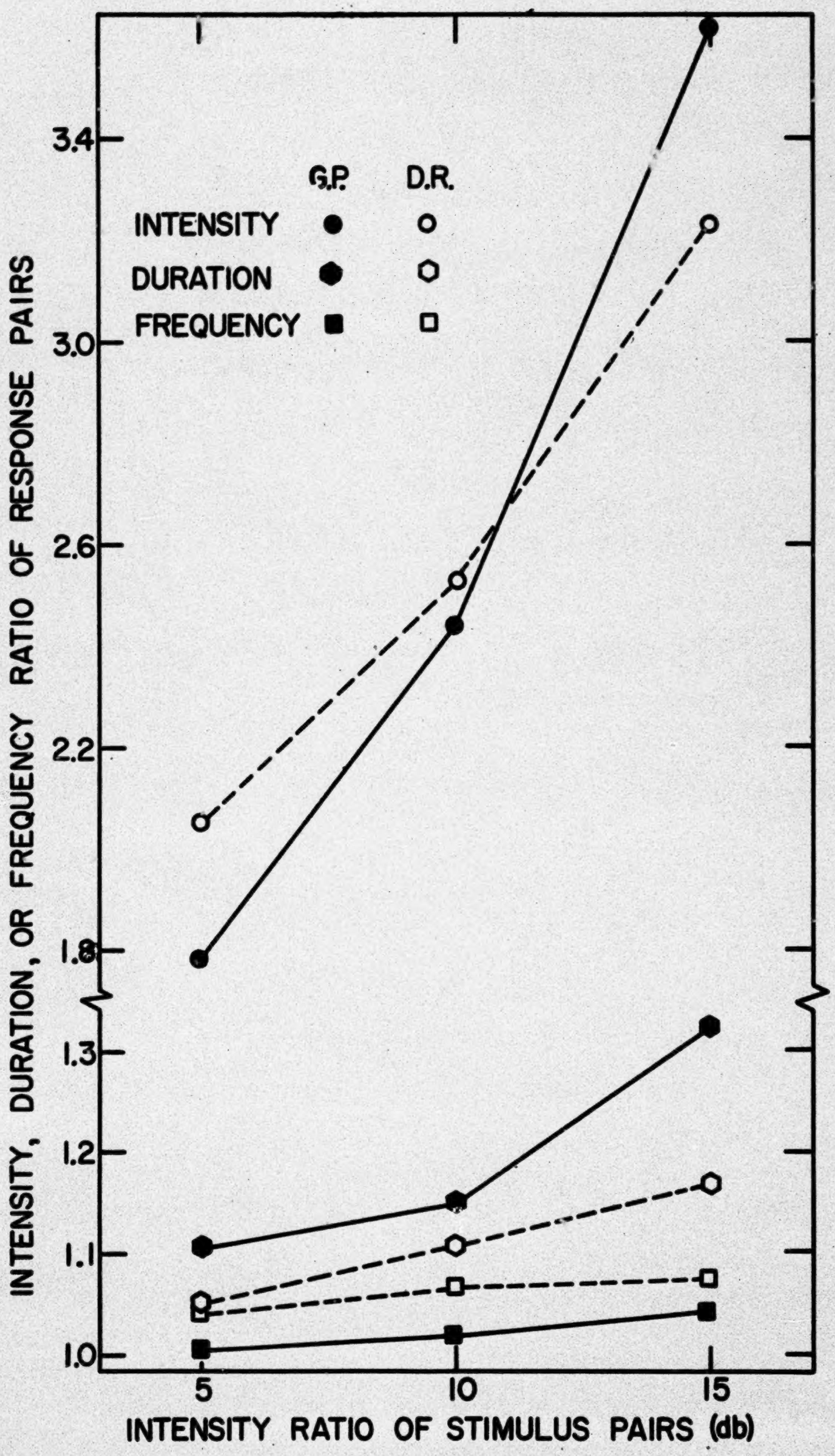
G.P.
D.R.
INTENSITY ● ○
DURATION ⬢ ⬡
FREQUENCY ■ □
INTENSITY, DURATION, OR FREQUENCY RATIO OF RESPONSE PAIRS
3.4
3.0
2.6
2.2
1.8
1.3
1.2
1.1
1.0
5
10
15
INTENSITY RATIO OF STIMULUS PAIRS (db)

Fig. 2

OPERANT RECONDITIONING OF A CONSONANT DISCRIMINATION IN AN APHASIC

H. L. Lane and D. J. Moore
The University of Michigan

OPERANT RECONDITIONING OF A CONSONANT DISCRIMINATION IN AN APHASIC

H. L. Lane and D. J. Moore[1]
The University of Michigan

The development of instrumentation and techniques for the acoustic analysis and synthesis of speech has made it possible to describe and control auditory discrimination of speech patterns with considerable precision. Research in acoustic phonetics is yielding an inventory of these discriminations in normal adult humans from a variety of linguistic communities (Fischer-Jørgensen, 1958). A few psychophysical studies have clarified some of the relations between discrimination of complex speech stimuli and the discrimination of simpler acoustic stimuli, varying only in one or two dimensions (Lane, Catania and Stevens, 1961; Fant, 1958).

These investigations have yielded a detailed description of speech discriminations in their "steady-state," and have led to a number of inferences concerning the "typical" course of speech discrimination learning in the ill-controlled and ill-controlling verbal community. However, there are no studies, to my knowledge, that have undertaken an experimental analysis of the acquisition of speech discriminations;[2] there are none employing operant conditioning techniques. At present, an account of the acquisition of speech discriminations must be founded on the results of a more rigorous inquiry into other discriminative behaviors. An experimental analysis of speech discrimination learning should show to what extent this extrapolation is valid. In the absence of this evidence, those concerned with the applied problem of manipulating speech discrimination, such as language teachers and speech therapists, have been unwilling to take the extrapolative leap. As a result, modern techniques for the control of behavior are rarely employed in these situations, where they might prove highly effective.

Armed with newly-won techniques for the control of the stimulus, the present study set out to condition a discrimination between /d/ and /t/ in an aphasic subject who was observed not to respond differentially to these stimuli when they were presented in isolation or in identical contexts ("minimal pairs"). The observed properties of this discrimination during acquisition and steady-state are relevant to the basic problem of the acquisition of speech discriminations as well as the applied problem of reconditioning in aphasia.

Method

Subject.—The following comments were extracted from the aphasic subject's medical record at The University of Michigan Hospital: "Age, 51 years. Medical diagnosis, cerebrovascular accident involving blood supply of the left middle cerebral artery, onset 1958; abnormal EEG focus in the left hemisphere (fronto-temporal), basic frequencies normal in both hemispheres; handedness, originally right, at present left; speech diagnosis, expressive-receptive aphasia and apraxia, severe dysarthria and dysrhythmia;[3] hearing, normal (pure audiometry)."

Speech stimuli.—The speech stimuli for this experiment were prepared by

A. M. Liberman at the Haskins Laboratories, using the Pattern Playback[4] to convert hand-painted spectrograms into sound. The spectrographic patterns employed, shown in Fig. 1, were identical except for the relative onset time of their first and second formants: the first formant was "cut back" in ten millisecond steps from 0 to 60 msecs. Liberman _et_ _al_. (1961) have shown that, with normal adults, the stimuli in this series evoke "labelling" responses that vary from /do/ for stimulus 0 to /to/ for stimulus 60. The non-speech, control patterns (Fig. 1, bottom row) were made by inverting the speech patterns, thus preserving the temporal variable while precluding speech recognition.

Order of stimulus presentation and procedure.—In order to measure the probabilities of /do/ and /to/ responses to the seven speech stimuli before and after training, these stimuli were first recorded on magnetic tape and later presented to the subject. The preparation of the stimulus tape has been described elsewhere (Liberman _et_ _al_., 1961):

> "[The stimuli were grouped in sets of three.] The various triads were made by pairing each stimulus with another stimulus having an onset delay that differed in the amount of 10, 20, or 30 msec. These pairs formed one-, two-, and three-step intervals respectively. Thus, for the one-step intervals, stimulus 0 was paired with stimulus 10, stimulus 10 with stimulus 20, etc. The two-step intervals were formed by pairing stimulus 0 with stimulus 20, stimulus 10 with stimulus 30, etc. The three-step intervals were made in similar fashion. Since there were seven stimuli (in each of the speech and control sets)there were six, five, and four comparisons in the 1-, 2-, and 3-step series respectively. Each stimulus comparison was arranged into four ABX (ABA, ABB, BAB, BAA) permutations for a total of sixty ABX triads (15 stimulus comparisons time four ABX permutations). Recorded copies of the 60 triads were made and distributed among eight tape sections of fifteen triads each. The fifteen triads were ordered randomly with the restriction that one and only one of the four ABX permutations be represented in each tape section. The tape sections were prepared with 0.5 msec. between members of the stimulus triad and 4.0 sec. spearation between triads."

An identical tape (C) was prepared comprised only of control stimuli. A third tape (T) was prepared for discrimination training, comprised only stimuli 0 and 60, copied from the stimulus tape described above. The training tape contained five sections:

1. Ten presentations of stimulus 0 followed by ten presentations of stimulus 60: 10 "0", 10 "60'
2. 5 "0", 5 "60', 5 "0", 5 "60'
3. 2 "0", 2 "60", 2 "0", 2 "60", 2 "0", 2 "60", 2 "0", 2 "60", 2 "0", 2 "60'.
4. 0-60-60-60--0-60-60-0-0-60-60-0-0-60-0-0-60-0-0-0--60-0-60-0-60-60.
5. 0-60-0, 0-60-0, 60-0-60, 0-60-60, 60-0-0, 0-60-0, 0-60-0, 60-0-60, 0-60-60, 60-0-0.

In sections 1, 2, and 3 the stimuli were separated by 4 seconds; in section four by 2 seconds, in section 5 by 0.5 secs. within triads and 4.0 secs. between triads.

Table I shows the order of presentation of the speech, control, and training tapes to the subject on various days over a total period of 48 days. On days one through nine, the experiment was conducted in a small, reverberant classroom. On days 35 through 48 it was conducted in an office. The appropriate tape was played at 7-1/2 ips on an Ampex 350 tape recorder with associated high-fidelity playback amplifier and loudspeaker. The subject was seated at a desk, facing the loudspeaker, at a distance of five feet. The experimenter sat in plain view adjacent to the tape recorder. On day one, the first tape to be presented was composed of the speech stimuli. The instructions to the subject were as follows:

> In this first part of the experiment you will hear a series of tones, in groups of three, with a few seconds between each group. We'll call the first tone in each group "Tone A," the second "Tone B," and the third "Tone C." Now, these tones are made by a machine, but they will approximate human speech sounds. The tones will be one-syllable sounds and will sound like /do/ or like /to/. Your task will be to darken in the space under the letter for each tone in each group which sounds like /do/.
>
> Don't worry about any trouble you may have in making discriminations. Different people hear the tones differently. Just try to decide whether the sound you hear is more like /do/ or more like /to/. If tone "A" is more like /do/, put a mark under "A" between these dotted lines [pointing]. If it is more like /to/, don't put anything. Listen to tones "B" and "C" for each group in the same way. All three tones may sound alike, all may be different, or any combination may occur.
>
> There is no right or wrong answer. You just listen to each tone and mark it down if it sounds like /do/; don't mark it down if it sounds like /to/.

After this tape had been presented, it was rewound and presented once again, with a different set of instructions:

> "In this second run through the tape, we'll be doing something a little different. This time, don't worry about whether the tones sound like /do/ or /to/, but just listen to them and try to remember what they sound like.
>
> As before, you'll hear the tones in groups of three. Tones "A" and "B" will differ from one another, sometimes very much, sometimes hardly at all. Listen to them and try to remember what they sound like. Tone "C" in each group will be exactly the same as either "A" or "B". So the third tone will sound like either the first or the second. If tone "C" sounds to you like tone "A", mark the space under "A". If it sounds like tone "B", mark the space "B". You won't have to make any mark under "C".
>
> Just as it was sometimes difficult for you to tell whether the tones sounded like /do/ or /to/, here it will be hard to tell whether the third tone is more like the first or the second. But don't worry... just listen to the tones and if the third tone sounds like "A" put a mark under "A". If it sounds like "B" to you, put a mark under "B".

Use anything about the word or sound that you want to use to tell whether the third sound is like the first or the second. If you are uncertain, guess. Always put something down."

These latter instructions defined an "ABX procedure" widely used in psychophysical experiments to measure the difference limen (Stevens, 1958). Liberman *et al*. (1961) have shown that with speech stimuli the two procedures, labelling and ABX, sample the same discriminative repertory: response probabilities obtained with the one method may be predicted accurately from those obtained with the other. The tape composed of control stimuli was presented next on day one with identical ABX instructions.

When S returned eight days later he was again given the ABX instructions but this time the stimulus triads were composed only of 200 and 2,000 cps tones. Four triads of the form ABA, ABB, BAA, and BAB were presented; each stimulus lasted 0.5 secs; the interval between stimuli was 0.5 secs, between triads, 4 secs. This stimulus series, requiring a relatively gross acoustic discrimination, was presented in order to establish that the subject's aphasia and partial paralysis did not impede responding appropriately to the ABX instructions. On the same day, S was again presented with the speech stimuli from tape A, twice in succession: once with the ABX instructions and once with the labelling instructions. As the last procedure for day nine, the Seashore Test of Musical Ability (Seashore, 1956) was administered. The four sections of this test permit relatively gross determinations of the subject's difference limen for "pitch, tonal memory, timbre, and rhythm." Since the test is standardized, a rough comparison of the subject's scores with the national norms could be made.

Twenty-five days later S returned and was trained to emit different responses to stimuli 0 and 60. The arrangement of the apparatus employed for conditioning this discrimination is schematized in Fig. 2. The subject sat opposite the experimenter at a table; a cardboard partition prevented him from seeing the experimenter's movements. E and S had matching pairs of buttons; each pair had one button labelled "DO" and another labelled "TO". A simple circuit composed of four buttons, a battery, a light bulb and some wire was constructed to provide reinforcement for the appropriate response. The training procedure was as follows. It was first established that the experimenter (E) could reliably discriminate between stimulus 0 and stimulus 60. The training tape (T) described above was played to S and E alike. When stimulus 0 was presented, E pressed his "DO" button and held it depressed until the next stimulus. When stimulus 60 was presented, he depressed his "TO" button. If S pressed the same button depressed by E during the inter-stimulus interval, the light bulb would illuminate. This crude arrangement of equipment for discrimination training was permissible since E's reaction time was always shorter than S's. Therefore, reinforcement always followed immediately after S emitted the correct response.

On the next day, S was again presented with the stimuli of tape A; once with the labelling instructions and then again with the ABX instructions. Nine days later the subject was again instructed to label the stimuli of tape A. On the final, 48th day, the subject responded to tape A according to the ABX instructions.

This procedure provided eight determinations of response probabilities to stimuli along the /do/ - /to/ continuum employed, two determinations with labelling instructions and two with ABX instructions before training and a similar set after training. It also provided a measure of the subject's discrimination of comparable non-speech stimuli (the control patterns), his ability to follow the ABX instructions, and his score on the several subtests of the Seashore Test of Musical Ability.

Results and Discussion

Figure 3 shows the probability of a /do/ response to each of the seven speech stimuli on the two determinations before and after training. These probabilities are expressed as the ratio of the number of /do/ responses to the total number of responses (in per cent) for each stimulus. From the negative slope of the pre-training determinations, shown in Fig. 3, it is clear that there was a gradual increase in the tendency to emit a /to/ response as the delay in first formant onset of the speech stimulus was increased. There was, furthermore, an overall reduction in the frequency of /do/ responses from Trial 1 to Trial 2 prior to training, although this change does not seem to have been under stimulus control.

Figure 4 permits a contrast between the pre-training labelling by the aphasic subject and a normal subject. The normal subject's data were selected at random from a set (N = 45) obtained in a classroom replication of "day one" of the present experiment, but are representative of the entire set and comparable to the findings of Liberman _et_ _al_. (1961). Allowing for a difference in the aspect ratio of the two graphs, it is clear that the aphasic's data do not reveal the abrupt transition from /do/ to /to/ responses shown by the normal. Since both sets of data show a gradient of probability of response along the stimulus continuum, it would be misleading to assert that one subject "makes a phonemic contrast" while the other does not, or that "perception is categorical" in one case and not in the other. Inference from the data in Figs. 3 and 4 suggests that, at least in the case of consonants, phonemic contrast is a matter of the relative steepness of the generalization gradient, since the transition region in these gradients is the acoustic correlate of the phoneme boundary.

Figure 3 shows the effects of 15 minutes of discrimination training with the two extreme stimuli from the continuum. Before examining the effects of this training on the relative frequency of /do/ and /to/ responses, the behavior of the subject during training will be described. On each of three presentations of section 1 of the training tape (see Method) the subject labelled each of the ten replications of stimulus 0 /do/, the first presentation of stimulus 60 /do/, and the remaining nine replications of stimulus 60 /to/. In each of two presentations of section 2 of the training tape, S labelled the five replications of stimulus 0 /do/, the first presentation of stimulus 60 /do/ and the remaining four replications of stimulus 60 as /to/. (Additional data on perseveration are presented later.) During the third presentation of section 2, S labelled all occurrences of stimulus 0 /do/ and stimulus 60 /to/. This performance was sustained throughout sections 4 and 5 of the training tape.

Figure 3 shows that a minimum of discrimination training effected a marked change in the generalization gradient. (A later paper will deal with the abruptness of discrimination learning in human subjects.) A determination of response probabilities on the day following training (post-training, trial 1) showed a gradient extending from 88 per cent /do/ responses to stimulus 0, to 6 per cent /do/ responses to stimulus 60. It is particularly interesting to note that a retest ten days later (trial 2) revealed a further steepening of the generalization gradient without further training in the laboratory. It is not known what contingencies may have arisen in the interim between Trials 1 and 2; thus the observed "self-sharpening" of the discrimination may not be replicable. The observed change in discriminative behavior (Fig. 3) produced by training with two stimuli from a unidimensional continuum and then testing at several points along the continuum is comparable to that obtained with conditioning procedures employing other human and subhuman operants and other stimulus continua.

Figure 5 shows the change in ABX discrimination resulting from training. On the first determination prior to training, S almost invariably blacked in the "A" column on his answer sheet. Since all stimulus pairs, (with 1-, 2-, and 3-step differences in first formant onset) were presented in four types of triad (ABA, ABB, BAA, and BAB), responding exclusively with "A" yielded 50 per cent correct discrimination. Eight days after these results were obtained a test series (P) was administered, composed of 200 and 2,000 cps tones as the A and B stimuli in four triads: ABA, ABB, BAA, and BAB. In each case S responded correctly, indicating that he could follow the ABX instructions appropriately. The Seashore Test of Musical Ability was then administered to give a gross indication of how the subject compared with a normal population in making certain psychophysical judgments. The aphasic's scores on the subtests, expressed as percentile ranks in the normal population were as follows: pitch, 10th percentile; tonal memory,10th percentile; timbre, 10th percentile; rhythm, 6th percentile.

A second determination of ABX discrimination was then made with the speech stimuli. Once again, S gave only "A" responses, yielding chance levels of discrimination. The average per cent correct on the two pre-training trials is shown by the broken curve in Fig. 5. The aphasic's discriminative behavior may once again be compared to that of a normal subject (Fig. 4, right). In the 1-step comparisons, the number of correct responses does not differ significantly from chance. In the 2- and 3-step comparisons, however, it is clear that the normal subject could discriminate among the A and B stimuli of the triads. As Liberman _et al_. (1961) have shown, the per cent correct for each stimulus pair is related to the per cent difference in /do/ responses evoked by the component stimuli under the labelling instructions. Thus, the peaks in the 1-, 2-, and 3-step discrimination curves for the normal S correspond to stimulus pairs with large or maximal differences in the per cent /do/ responses evoked by each. Therefore, the chance levels of ABX discrimination shown by the aphasic prior to training (Fig. 5) were not unexpected in view of the small differences in the frequency of /do/ responses to the several stimuli prior to training (Fig. 3).

The solid lines in Fig. 5 show the per cent correct responses to the ABX triads in the two determinations following training. These data lend support to the suggestion (made in another form by Liberman et al.) that the labelling and ABX procedures sample the same discriminative repertory. The training procedure consisted solely of training two labelling responses to stimuli 0 and 60. The subsequent change in labelling behavior (Fig. 3) seems to be correlated with a change in correct discrimination, measured under ABX instructions (Fig. 5). Furthermore, the peak in the discrimination function for the 3-step comparisons occurs with A and B stimuli selected from opposite sides of the phoneme boundary. Figure 6 shows this graphically by presenting the combined pre- and post-training labelling and discrimination results for the aphasic subject. The peak in the 3-step discrimination function occurs with those A and B stimuli that were most disparate in the frequency of /do/ responses evoked during labelling.

Discrimination training not only altered the shape of the 3-step discrimination function, it also raised the overall frequency of correct responses above the chance levels that were observed prior to training. Liberman et al. (1961) have posed the question of whether speech discriminability functions reflect "acquired distinctiveness, acquired similarity, or both." In the case of normal human adults, where the experimenter has not participated in the conditioning process, this question can be answered only by observing the behavior of the subject under the control of comparable non-speech stimuli. If the speech stimuli are more discriminable than the controls at all points, acquired distinctiveness is observed. If they are less discriminable at all points, this is evidence for acquired similarity. An intermediate position of the speech function relative to the control indicates both, acquired similarity and acquired distinctiveness. The choice of a comparable set of non-speech control stimuli is extremely difficult, however.[5] Two mutually exclusive features are usually desired: (1) the control stimuli should be comparable to the speech stimuli so that behavior under each condition is strictly comparable; (2) the control stimuli should not evoke the speech discrimination under investigation, so that their discriminability may serve as a baseline for assessing the effects of training in speech discrimination. In the present study these two requirements have been met by employing the speech stimuli as the control stimuli. Prior to training, the speech stimuli evoked the speech discrimination under study only to a very slight extent, yet they were indeed identical to the stimuli employed after training to assess its effects.

The speech discrimination function shown in Fig. 6 reflects acquired distinctiveness exclusively. This is the result of the particular discrimination training procedure employed. Prior to training, ABX discrimination of the speech stimulus (tape A) and the control stimulus (tape C) did not exceed chance levels. Following training, speech discrimination was superior to control discrimination at all points.

Comparison of Figs. 4 and 6 shows that a limited amount of conditioning under controlled conditions went a long way toward developing normal phoneme discrimination. These findings are the more noteworthy in view of the age and medical condition of the subject. We may now inquire whether discrimination training had any effect on the subject's perseverative behavior, mentioned earlier. The perseverative tendency of aphasics has been widely cited and discussed (e.g., Tikofsky and Reynolds, 1961). Figure 7 shows the effect of discrimination training on the perseverative behavior of the aphasic subject in the present experiment. Prior to discrimination training (filled histograms), responses to the stimulus tape following ABX instructions were, as indicated earlier, highly stereotyped. On the first pre-training determination there was one instance of 108 identical responses ("A") in a row. Pre-training Trial 2 shows a slight decrease in length of runs of identical responses. Perseverative behavior following discrimination training is greatly reduced (shaded histograms). Perseveration is lower on Trial 2 than on Trial 1 in both pre- and post-training determinations. The frequency distribution of identical triads in a run for the aphasic after training more nearly approximates that distribution for the stimulus series (unfilled histograms). The correlated change in ABX and labelling behavior following discrimination training has been discussed earlier. The third change observed in the behavior of the aphasic, a reduction in perseveration, is related to the first two and, like them, is the consequence of discrimination training. The increase in correct ABX discrimination, shown in Fig. 6 could not have been obtained without a decrease in perseveration (although the converse is not true).

Summary

A technique for the synthesis of speech was employed to investigate the acquisition of a consonant discrimination by an aphasic subject. It has been shown by Liberman (1958) that a change of 10 to 20 msecs. in the relative onset times of the first and second formants in a particular spectrographic pattern was sufficient to shift the frequency of labelling responses to the acoustic correlate of the pattern from 75 per cent /do/ to 75 per cent /to/ with normal subjects. When this synthetic speech series was presented to an aphasic subject, there was only a very slight tendency for the frequency of /do/ responses to decrease as the first formant was "cut back" in six ten-millisecond steps. When these stimuli were arranged in triads and presented for ABX discrimination, only chance levels of correct responding were obtained. Despite an equal frequency of triads of the form ABA and ABB, BAA, and BAB, in the stimulus series, the aphasic subject repeatedly marked "A" on his answer sheet after the presentation of each triad.

The two stimuli from the extremes of the continuum (0 msecs. cutback and 60 msecs. cutback of the first formant) were then used to condition a discrimination. When stimulus 0 was presented and S pressed a button labelled /do/, he received a flash of light; when stimulus 60 was presented, pressing an alternate button labelled /to/ was similarly reinforced. Differential responding was obtained after a few minutes of conditioning. A dramatic change in the discriminative responses of S to the stimulus continuum was then observed. The generalization gradient obtained after training was much steeper and approximated that found with normal subjects. Correct discrimination of stimulus pairs differing by 30 msecs. in the amount of first formant cutback exceeded chance levels for all pairs. The frequency distribution of perseverative responses after training approximated that distribution of identical triads in the stimulus series.

The discriminative behavior observed following training of the aphasic subject was compared to that observed in normal adults. Inferences concerning the acquisition of native phoneme discriminations by normal subjects were assessed in the light of the training procedures and their effects in the present study.

FOOTNOTES

1. This research was conducted at the Communication Sciences Laboratory, The University of Michigan. The interest and assistance of Mr. A. Zoss, Dr. R. Tikofsky, and the Haskins Laboratories are gratefully acknowledged.

2. This type of analysis requires that the experimenter systematically vary the stimulus members of the "three term relation" discriminative stimulus, response, reinforcement, discussed by Skinner (1957).

3. For a definition of terms, see Wepmen, 1951.

4. For a description of the Pattern Playback and its use see Cooper *et al.*, 1951.

5. Cf. discussion by Liberman *et al.* (1961).

REFERENCES

Cooper, F. S., Liberman, A. M., and Borst, J. The inter conversion of audible and visible patterns as a basis for research in the perception of speech. *Proc. Nat. Acad. Sci.*, 1951, 37, 318-325.

Fant, C. G. M. Modern instruments and methods for acoustic studies of speech. *Proc. VIII Int. Congress of Linguists*. Oslo: Oslo University Press, 1958.

Fischer-Jørgensen, E. What can the new techniques of acoustic phonetics contribute to linguistics? *Proc. VIII Int. Congress of Linguists*. Oslo: Oslo University Press, 1958.

Lane, H. L., Catania, A. C. and Stevens, S. S. Voice Level: Autophonic scale, perceived loudness, and effects of sidetone. *J. Acoust. Soc. Amer.*, 1961, 33, 160-167.

Liberman, A. M., Delattre, T. C., and Cooper, F. S. Some cues for the distinction between voiced and voiceless stops in initial position. *Lang. Speech*, 1958, 1, 153-157.

Liberman, A. M., Harris, K. S., Kinney, J. A. and Lane, H. L. The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *J. exp. Psychol.*, 1961, 61, 379-388.

Seashore, C. E. *Seashore measures of musical talent*. Revised Edition. Grades 4-16 and adult. New York: The Psychological Corporation, 1956.

Skinner, B. F. *Verbal Behavior*. New York: Appleton-Century-Crofts, 1957.

Stevens, S. S. Problems and methods of psychophysics. *Psychol. Bull.*, 1958, 54, 178-196.

Tikofsky, R. and Reynolds, G. L. Nonverbal learning and aphasia. *J. Speech Hear. Disord.* (in press).

Wepman, J. M. *Recovery from Aphasia*. New York: Ronald Press, 1951.

TABLE I

ORDER OF PRESENTATION OF STIMULUS TAPES

Day	Tape (description)		Response mode	Time (mins.)
1	A	speech stimuli	pencil mark* to each /do/	40
1	A	speech stimuli, repeat	pencil mark* under "A" or "B"	40
1	C	control stimuli	pencil mark* under "A" or "B"	40
9	P	200 and 2,000 cps tones	pencil mark* under "A" or "B"	15
9	A	speech stimuli	pencil mark* under "A" or "B"	40
9	A	speech stimuli, repeat	pencil mark* to each /do/	40
9	MA	Seashore test of musical ability	write S (same) or D (different) for each stimulus pair	30
35	T	training: stimuli 0 and 60	button press (see text)	15
36	A	speech stimuli	pencil mark* to each /do/	40
36	A	speech stimuli, repeat	pencil mark* under "A" or "B"	40
46	A	speech stimuli	pencil mark* to each /do/	40
48	A	speech stimuli	pencil mark* under "A" or "B"	40

---

* On an IBM answer sheet.

FIGURE CAPTIONS

Fig. 1. Spectrographic patterns which were converted to sound by the Pattern Playback to form the stimuli of the experiment. (After Liberman, et al., 1961).

Fig. 2. Schematic of equipment arrangement for conditioning a /do/ - /to/ discrimination.

Fig. 3. Per cent /do/ responses to each of the seven stimuli before and after discrimination training. Eight days elapsed between Trial 1 and Trial 2 of the pre-training determinations. Nine days elapsed between Trials 1 and 2 of the post-training determinations. The number of phoneme labelling responses represented by each percentage point depends on the corresponding stimulus: 0, 36; 10, 48; 20, 60; 30, 72; 40, 60; 50, 48; 60, 36.

Fig. 4. At left: per cent /do/ responses to each of the seven stimuli before discrimination training by a normal adult subject. The number of phoneme labelling responses represented by each percentage point depends on the corresponding stimulus: stimulus 0, 36; 10, 48; 20, 60; 30, 72; 40, 60; 50, 48; 60, 36. At right: discrimination functions (ABX method) for the 1-, 2-, and 3-step differences among the synthetic speech stimuli. Each point represents the number of correct "A" or "B" responses x 100/24.

Fig. 5. Aphasic subject. Discrimination functions (ABX method) for the 1-, 2-, and 3-step differences among the synthetic speech stimuli. The two determinations before training were combined, as were the two following training. Each point represents the number of correct "A" or "B" responses x 100/16.

Fig. 6. Combined pre- and post-training labelling and ABX discrimination data for an aphasic subject.

Fig. 7. Perseverative behavior of an aphasic subject before and after discrimination training. The unfilled histograms show the number of occasions in the stimulus series on which there were two, four, and five identical triads (of the form ABA, ABB, BAA, or BAB) in succession. The shaded histograms show the number of such "runs" for the aphasic subject before training, while the filled histograms show the number of runs after training.

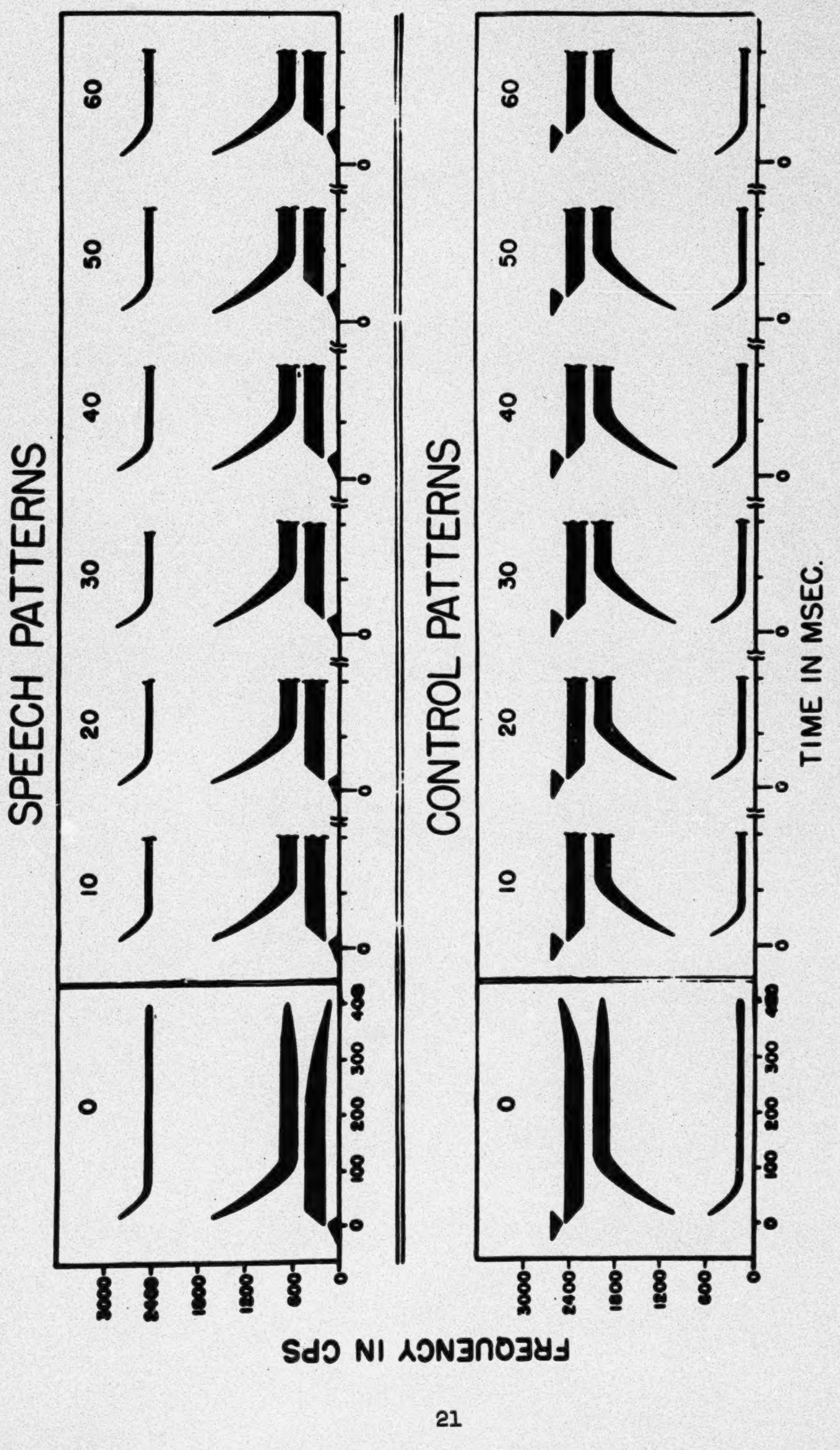
SPEECH PATTERNS
CONTROL PATTERNS
0
10
20
30
40
50
60
FREQUENCY IN CPS
3000
2400
1800
1200
600
0
0
100
200
300
400
TIME IN MSEC.

Fig. 1

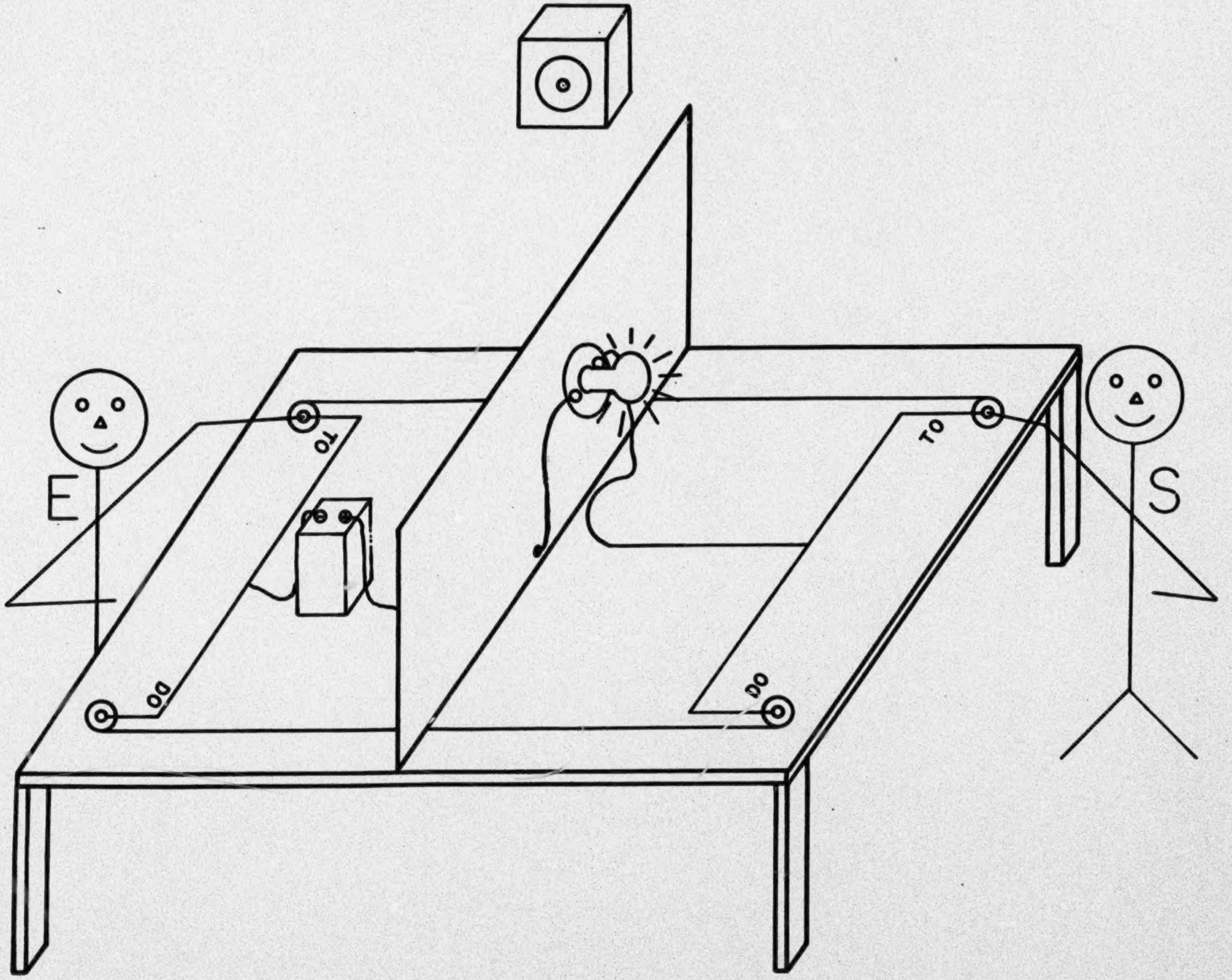
E
S
TO
DO
TO
DO

Fig. 2

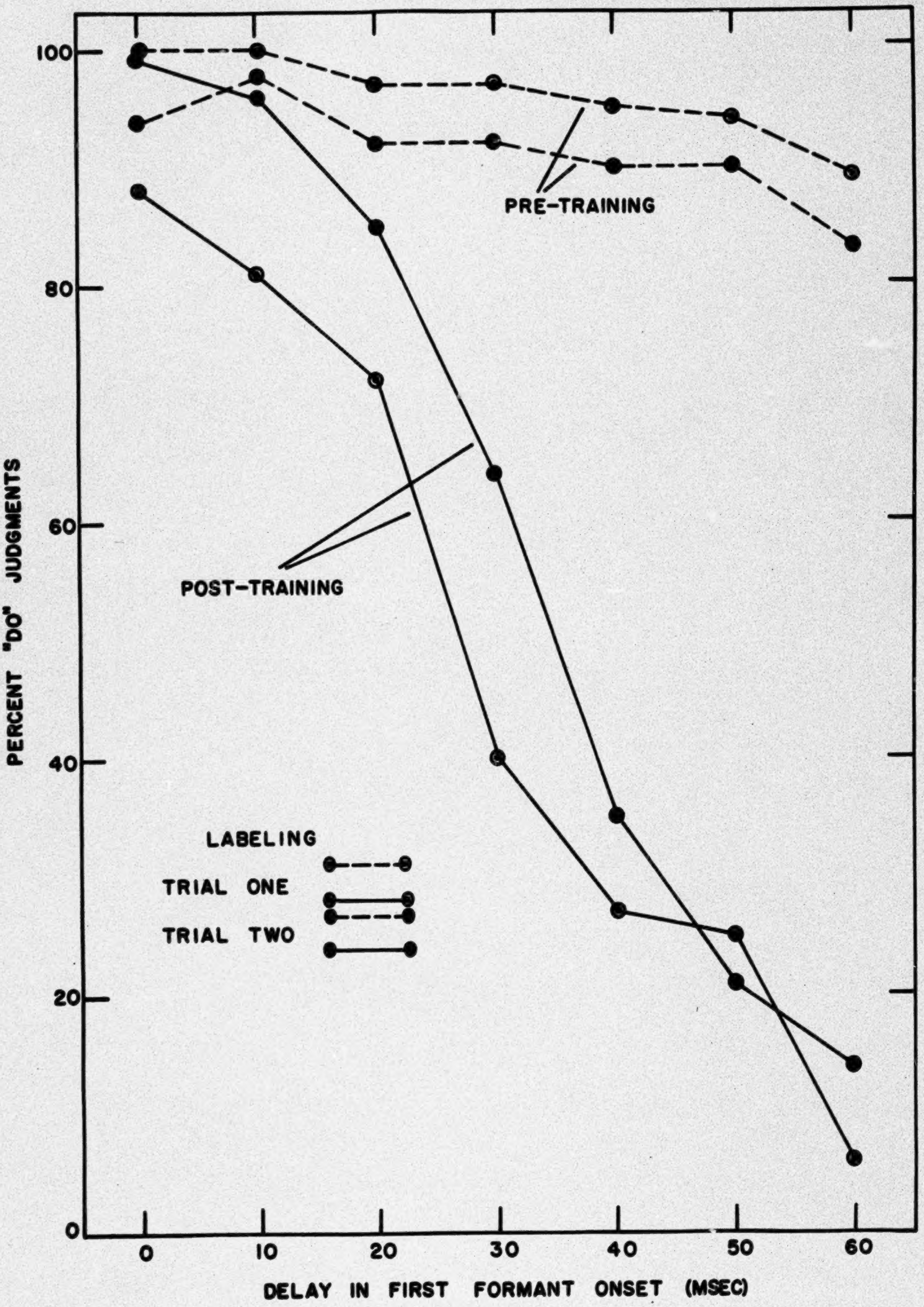
PERCENT "DO" JUDGMENTS
100
80
60
40
20
0
0
10
20
30
40
50
60
DELAY IN FIRST FORMANT ONSET (MSEC)
PRE-TRAINING
POST-TRAINING
LABELING
TRIAL ONE
TRIAL TWO

Fig. 3

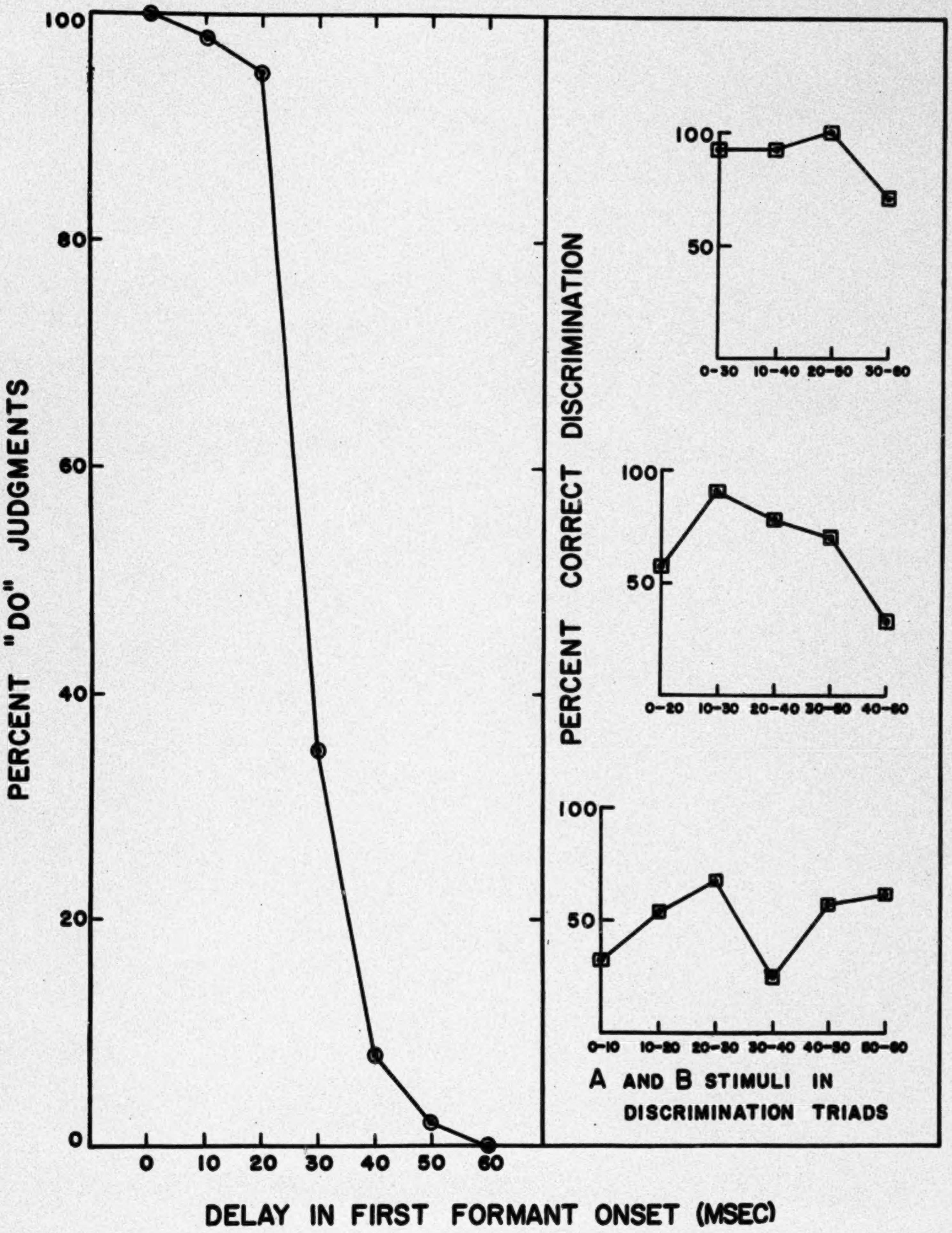
PERCENT "DO" JUDGMENTS
100
80
60
40
20
0
0
10
20
30
40
50
60
DELAY IN FIRST FORMANT ONSET (MSEC)
PERCENT CORRECT DISCRIMINATION
100
50
0-30
10-40
20-50
30-60
100
50
0-20
10-30
20-40
30-50
40-60
100
50
0-10
10-20
20-30
30-40
40-50
50-60
A AND B STIMULI IN DISCRIMINATION TRIADS

Fig. 4

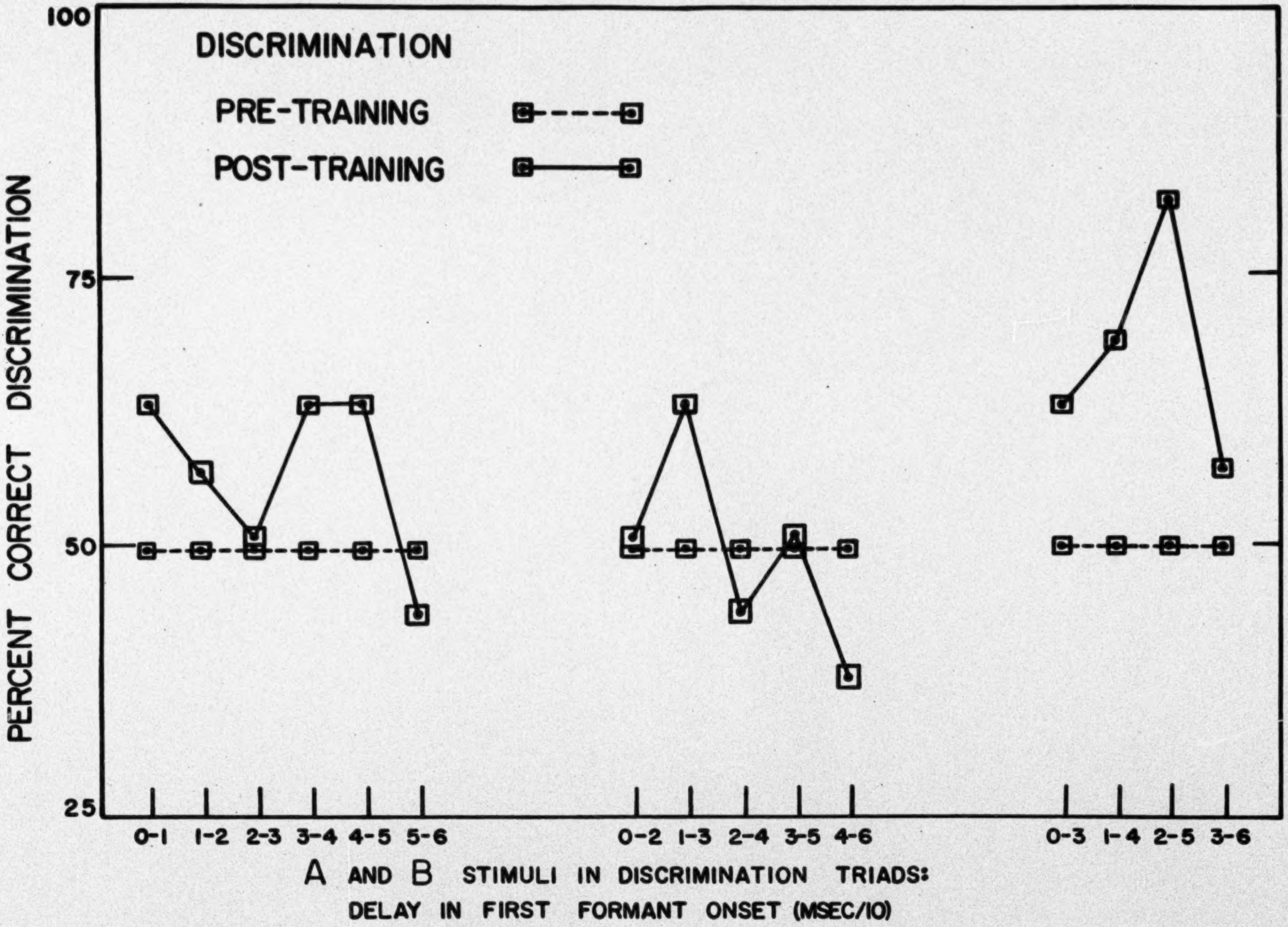
PERCENT CORRECT DISCRIMINATION
100
75
50
25
DISCRIMINATION
PRE-TRAINING
POST-TRAINING
0-1
1-2
2-3
3-4
4-5
5-6
0-2
1-3
2-4
3-5
4-6
0-3
1-4
2-5
3-6
A AND B STIMULI IN DISCRIMINATION TRIADS:
DELAY IN FIRST FORMANT ONSET (MSEC/10)

Fig. 5

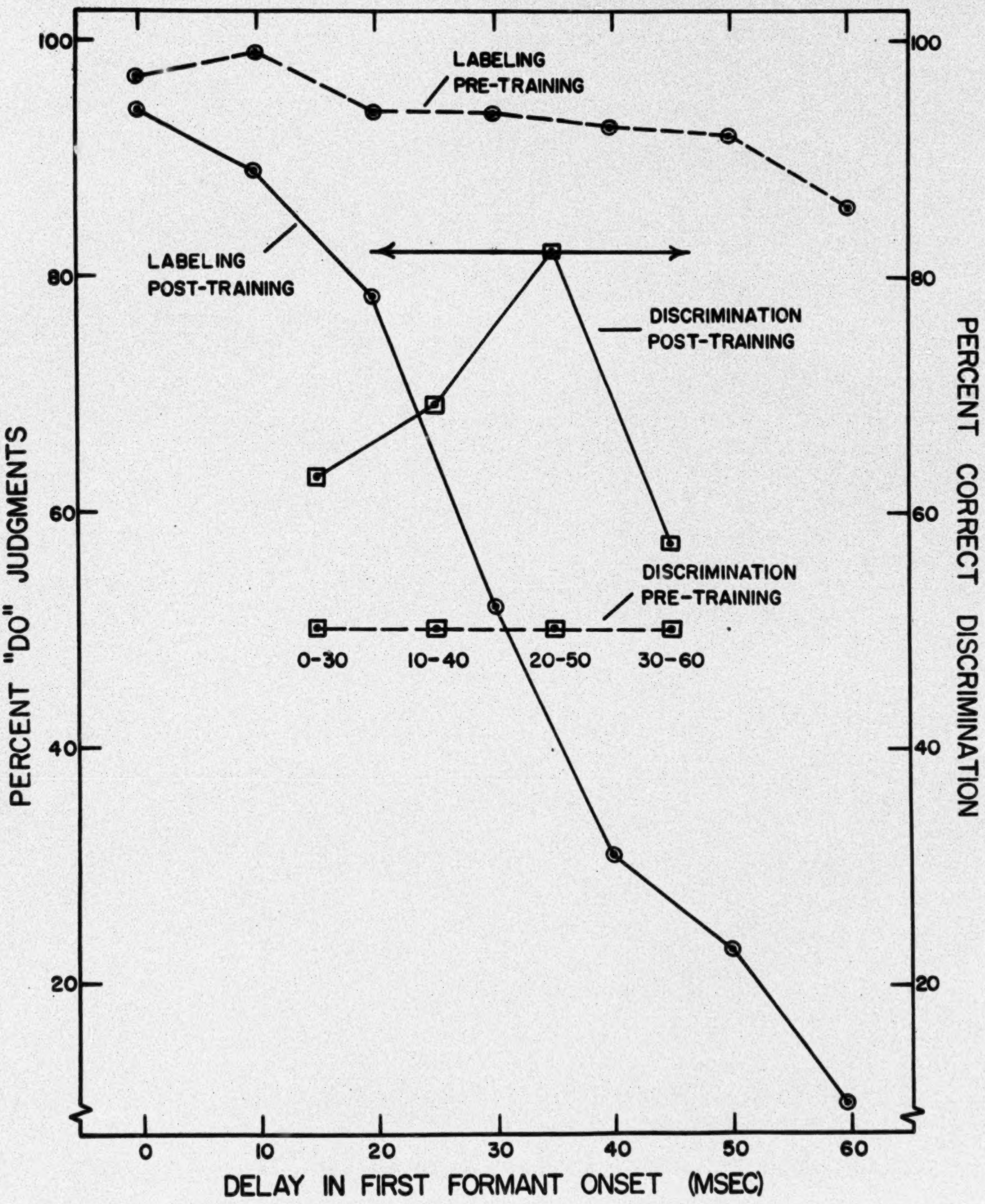
LABELING
PRE-TRAINING
LABELING
POST-TRAINING
DISCRIMINATION
POST-TRAINING
DISCRIMINATION
PRE-TRAINING
0-30
10-40
20-50
30-60
100
80
60
40
20
0
10
20
30
40
50
60
PERCENT "DO" JUDGMENTS
PERCENT CORRECT DISCRIMINATION
DELAY IN FIRST FORMANT ONSET (MSEC)

Fig. 6

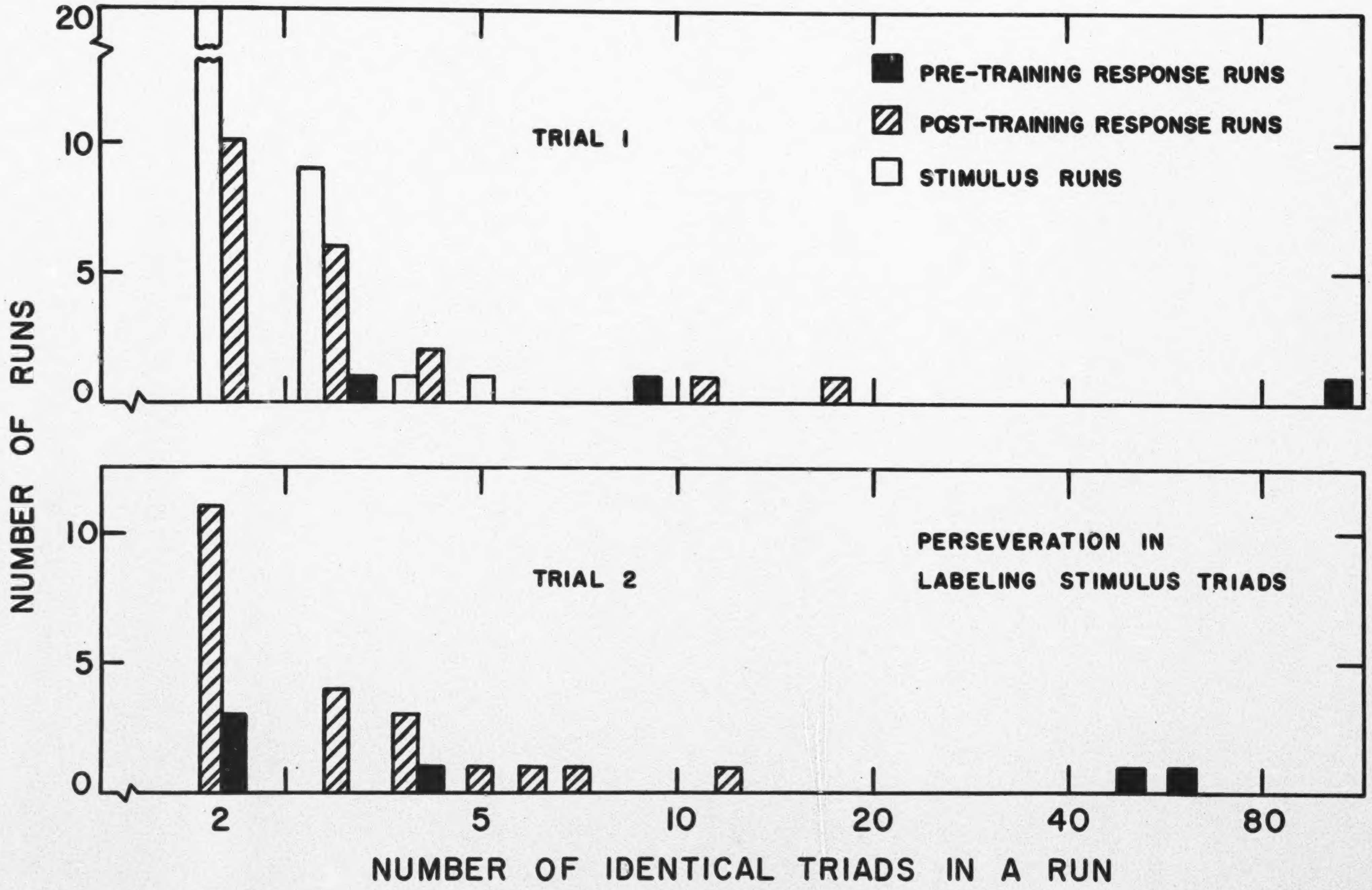
20
10
5
0
NUMBER OF RUNS
TRIAL I
■ PRE-TRAINING RESPONSE RUNS
▨ POST-TRAINING RESPONSE RUNS
☐ STIMULUS RUNS
10
5
0
TRIAL 2
PERSEVERATION IN
LABELING STIMULUS TRIADS
2
5
10
20
40
80
NUMBER OF IDENTICAL TRIADS IN A RUN

Fig. 7

TEACHING MACHINES AND PROGRAMMED LEARNING

Harlan Lane
The University of Michigan

TEACHING MACHINES AND PROGRAMMED LEARNING

Lumsdaine, A. A. and Glaser, R., Department of Audio-Visual Instruction
National Education Association, Washington D.C., 1960. 724 pp. $7.50

Harlan Lane
The University of Michigan

The express purpose of "Teaching Machines and Programmed Learning" is "to provide a comprehensive reference source on teaching machines and the techniques of instruction that are associated with them" (p. 1). It is unclear what role "a comprehensive reference source" will play in the new era that this book heralds, an age of educational technology. When behavior and the environment are engineered to specifications and learning is, by definition, programmed learning, the student found reading a comprehensive reference source may be accused of non-adaptive sentimentalism akin to serving tea and ices during weightless flight; presumably, he would be appropriately reconditioned.

However uncertain the book's future role in the marvelous world it forecasts, its present role is clear: it serves both as gadfly and as guide for the modern educator. Among the 47 papers by distinguished psychologists, educators, and engineers collected here, we find reports of the discovery of a science of human behavior and the prospects for the utilization of this science to change the condition of man. Needless to say, a change in the condition of man means a change in man himself. Few who have read the book would say that the issues in question are of less moment than its occasionally messianic tone implies.

In part I of "Teaching Machines and Programmed Learning" the editors describe the purpose and scope of the book by presenting a "review," "overview," and "preview" of the field. In part II, S. L. Pressey and his co-workers describe some early attempts to construct test scoring devices that would also have value for self-instruction. Part III of the book presents selected writings and research by B. F. Skinner and his students. Here, the reader may view the fruits of applying the principles of the experimental analysis of behavior to education; he may also glimpse the strategy and tactics of the science of behavior on which these applications are based. Four completed programs and their associated devices are described, along with initial findings obtained in the school setting. Additional concepts, programs, devices, and extensive "field tests" are described in a later section (infra).

Test scoring devices and the experimental analysis of behavior are only two of the points of departure for writers and researchers in the field of automated teaching. Part IV describes another starting point: specific training needs. The articles in this section of "Teaching Machines and Programmed Learning" are characterized by a greater diversity of equipment and techniques and by a greater emphasis on the acquisition of nonverbal skills than is apparent in part III.

Part V provides further evidence that the application of behavioral science to education is not only an achievement devoutly to be wished for but also a present reality. Here are presented the results of recent experiments in laboratories and schools along with an examination of the implications of these findings.

Two appendices complete this major work: "Appendix I is an annotated compilation of papers in the field of teaching machines and programmed learning. Appendix II is a consolidated bibliography of all the references [cited in the book] " (p. 574).

The Identity of Teaching Machines and Programmed Learning

While the purpose of the authors of "Teaching Machines and Programmed Learning" may have been to provide a comprehensive reference source, they have accomplished something much more significant and far-reaching. The present collection of articles has defined teaching machines and programmed learning by colligation. Through the contributions of the 47 authors, programmed learning has taken on an identity --- an identity that is misleading, inconsistent, self-contradictory; an identity that cannot but militate against the long-term efficacy of programmed learning. Teaching machines and programmed learning are at once identified with (1) Socrates, (2) aids to education, including audio-visual aids, self-scoring devices, and computers, and (3) behavioral science. There is no doubt that Socrates and aids to education have adventitious properties in common with teaching machines and programmed learning. However, to say that a test-scoring device, for example, is a teaching machine --- or to include a description of such devices in a text on teaching machines --- is to engage in metaphor. The metaphor is understandable in the light of the uncertain identity of programmed learning. As Skinner has pointed out, "In a novel situation to which no generic term can be extended, the only effective behavior may be metaphorical." However, "scientific verbal behavior is set up and maintained because of certain practical consequences;" metaphor cannot serve science well (Skinner, 1957). In view of the growing scientific, social, and commercial interest in programmed learning its proper identification should be considered carefully.

Programmed Learning and Socrates

The identification of programmed learning with the Socratic method has several sources of strength (vide p. 5). In academic circles it is fashionable to view each advance in the humanities as a footnote to Plato or Aristotle. Perhaps an amusing comment on progress in education is also implied. Finally, the Socratic method does have a few features in common with programmed learning. However, programmed learning should not limit itself to the techniques of behavioral control exercised by Socrates and should not be identified with these techniques. The nature of the behavioral control that can be exerted by teaching machines far exceeds the powers of the ancient Greek. For example, Pask (p. 336) describes an electronic keyboard teaching machine that makes adaptive changes in the program based on error distributions and response latencies. A second example is provided by a device, designed by the reviewer, to teach prosodic features of speech, that makes adaptive changes in the program as it analyzes the mismatch in relative amplitude, fundamental frequency, and tempo of the stimulus and the echoic response of the subject.

Few teaching machines or programs presently utilize the true potential of automation. This potential will not be realized as long as machines and programs are viewed as so many private tutors.

Programmed Learning and Aids to Education

The name AIDS (auto-instructional devices) has been proposed for teaching machines; a more unfortunate choice of name could hardly have been made. Properly designed teaching machines, along with their programs, are not aids at all but teacher surrogates for the behaviors that they develop. In accordance with Porter's classification of teaching aids and devices (pp. 116, 117), it is proposed that "aids" be reserved for those techniques or equipments "which must be supplemented by some means, usually a teacher, in order to be effective" (p. 118).

The identification of teaching machines and programmed learning with aids to education such as movies, self-scoring devices, and computers, rather than with a science of behavior, has let to unfortunate inconsistencies in approaches to the improvement of education. The new technology of education described so vividly by Ramo (p. 367), for example, shows a great deal of sophistication in the presentation of stimuli and in the processing of behavioral data, while evidencing little or no sophistication in the modification of the behavior that links the two and is, after all, the goal of the entire process. In a wondrous world of automatic student recognition, automatic curriculum selection, and automatic performance analysis, it verges on the comic to read that: "...the student is allowed a period for undisturbed contemplative thought before registering his answer" (p. 373). Every step in the educational process that Ramo describes is engineered to specifications except the behavior itself! Yet the possibilities of behavioral engineering are as great, and the potential profits as many, as those derived from electronic engineering. "Educational technology" may become an oxymoron if it denotes an admixture of the marvel that is electronics and the anachronism that is educational practice. Instead, a conception of education is required that is consistent with our conception of other areas of applied science. The traditional image of man is a cartoon against the backdrop of modern science. We need the courage to draw up specifications for an educated man that are not specifications for ourselves and the willingness to control behavior to bring that man about.

The identification of teaching machines and programmed learning with aids to education obscures the true nature of the decision which the educator must make: a considered decision to adopt the materials and techniques of programmed learning implies an acceptance of a scientific conception of human behavior. Programmed learning will make only a slight fraction of its potential contribution to education if it is viewed only as an aid that will leave the teacher free for "developing [sic] in her pupils fine enthusiasms, clear thinking, and high ideals" (Pressey, p. 40). We must review the goals of education, specify the desired behaviors, and examine the means of obtaining --- not "developing" --- these behaviors, in the light of a science of behavior. To do less is dishonest.

The failure to identify programmed learning with a science of human behavior and the concomitant failure to appreciate its implications has let to a proliferation of devices, under the pressure of commercial profit-mongering, and to a willy-nilly trading of behavioral specifications for considerations of profit in machine design. There is more than a coincidental resemblance between the products of teaching machine manufacturers before and after they entered the field. The commercial pattern seems to be: (1) recognition of a potential market; (2) design of a prototype device based primarily on current production facilities and sales outlets; (3) consultation with psychologists, educators, or others to select the prevailing point of view that best validates the device designed in step (2); (4) preparation of literature and initial production run. Pressey seems to have pioneered this approach, fitting the theory to the device, when he said of his self-scoring apparatus, exhibited in 1924: "The somewhat astounding way in which the functioning of the apparatus seems to fit in with the so-called 'laws of learning' deserves mention in this connection" (p. 37). (The author goes on to enumerate the laws that, in retrospect, "fit in.")

The "products" of this approach range from several thousand dollar stimulus-presentation devices to fifty cent "sit and spit" test scoring devices (digital application of saliva to a treated card reveals which multiple choice letter, A, B, C, or D, is correct). Each of these miscreants masquerades under the topical heading of teaching machines with such magical names as the Didak 101, the Mentor, etc. The sales techniques employed make the Hidden Persuaders seem forthright and candid by invidious comparison.

The reviewer regrets to write that "Teaching Machines and Programmed Learning," far from ameliorating this situation, may be expected to aggravate it. Part IV of the book demonstrates the scope of application, occasionally proven, mainly potential, of teaching machines. The variety of devices and approaches presented here would be salutory were it not for the fact that, as it turns out, each author with a device considers himself a knowledgeable, however unique, behavioral scientist. For example, an article by Crowder in part IV of the book suggests several basic assumptions concerning human learning along with an underlying model whose appropriateness may well be questioned (cf. Glaser, p. 437). Thus, "we approach the design of a teaching machine as a problem in communication" (p. 298). "The primary purpose is to determine whether the communication was successful, in order that corrective steps may be taken by the machine if the communication process has failed" (p. 288). Crowder denies access to any "educational philosopher's stone" (p. 287); this "machine philosopher's stone" seems a poor substitute, however.

The contributors to part IV of "Teaching Machines and Programmed Learning" seem to have concluded that, since no one point of view is held unanimously among psychologists, any point of view is equally tenable. That this is obviously untrue is testified to by the superficiality and inconsistencies of the various behavior "theories" that abound in part IV. An extension of this logic, which the reader may well make, permits the educator to adopt those teaching machines, and those features of machines, that appear consistent with his personal philosophy. Such an outcome would be disastrous for the ultimate efficacy of automated teaching. What is required of the educator, on the contrary, is a re-evaluation of personal philosophy in the light of the principles of behavior that underlie the development and format of the technological revolution to which this text is testimony.

Programmed Learning and Behavioral Science

Unlike other recent changes in educational technology, the growing utilization of programmed materials has a surprising by-product that strengthens the very movement itself. It cannot be said of educational TV, for example, that its use in the school has led perforce to a wider understanding of electronics. However, a growing interest in programmed learning has led to an increasing awareness of the principles of behavior on which it is based. This is well illustrated in Barlow's report on the self-instruction program at Earlham College: "Each programmer so far has himself worked through at least a portion of a Holland-Skinner program for the natural science psychology course at Harvard. The programmers thus learn some of the background of the basic principles we are currently attempting to follow at the same time that they become familiar with the oldest program available" (p. 419). Barlow's rationale has proven equally appealing to many other psychologists and educators throughout the country; the Holland-Skinner program is widely used not only in introductory courses in psychology and education but also in advanced seminars. The recent paperback edition of the program should abet this development (Holland and Skinner, 1961).

The Holland-Skinner program has been, therefore, an important step toward identifying programmed learning with its parent discipline. The second major step in this direction is the collection of articles presented in part III of "Teaching Machines and Programmed Learning." This section of the book should go far in correcting the widespread misunderstanding of the relevance of laboratory research with humans and subhumans to problems in education. A comment by Mr. Crowder, "I have no quarrel with Skinner; when a man wants to have some pigeons trained I send him to Skinner" was received with great enthusiasm at a recent convention of the Department of Audio-Visual Instruction, NEA. It is appropriate, therefore, that this very organization should sponsor the publication of articles that may remedy this misunderstanding.

The concept that links the knowledge gained in the laboratory to its application in education is control. The recent advances in the science of learning have taken place because "the law of effect has been taken seriously; we have made sure that effects do occur and that they occur under conditions which are optimal for producing the changes called learning" (Skinner, pp. 99, 100). Effects do occur reliably, promptly, and under optimal conditions only when the environment is controlled. To the extent that we sacrifice this control we impair and deflect the learning process.

Questions and Research for Teaching Machines and Programmed Learning

Not only programs but also programmers and books about programmed learning are filled with questions. A question is both an effective way of evoking the behavior of others and also an effective way of evoking our own verbal behavior. The following questions, taken from various pages of "Teaching Machines and Programmed Learning," are presented in order to (a) evoke verbal behavior on the part of the reader, (b) suggest further the nature of research and writing in this area, (c) indicate some of the unresolved problems in programmed learning discussed throughout the book.

1. Which is better: branching or linear programming?

2. Which is better: multiple choice or constructed response modes? Is implicit responding inferior to overt behavior in learning?

3. Is automatic response scoring preferable to self-scoring?

4. Is a cheat-proof feature in machine design important?

5. The reinforcing control exerted by candy, points, "going on to the next item," and "making the gadget work" have all been demonstrated. Which reinforcers should be employed?

6. Are multiple programs, branching, or adaptive programming important in the light of individual differences?

7. What subject matters do not "lend themselves" to programming?

8. What is the optimal length of frame, length of set, and length of program? In constructed response programming, what is the optimal length of response?

9. What is the optimal length of time for a student to work on a program in one sitting?

10. How should prompts be introduced and vanished? What amount or rate of prompting is optimal?

11. What error rate is optimal? Is an error-repeat feature important? How many correct responses to an item should be required before it is dropped out of the program?

12. What are the preferred sequencing logics? What is the optimal size of step?

13. What are the best ways to maintain student motivation?

14. Do the verbal knowledge, motor skills, and study habits acquired through programmed learning transfer to other performances?

In the opinion of the reviewer, questions like those enumerated above are not effective stimuli for the type of research that is needed in the area of programmed learning. At best these questions point to some of the variables that control the behavior of the student. Since the student's behavior at any point is a function of the complex interaction of all these variables, and many others not cited, it is not possible to give a general answer to any single question nor, of course, to answer all at once.

Questions of the type: which is better, A or B? lead to a type of inquiry which we may call comparison research. This kind of research has an extensive tradition in education and psychology and its pursuit probably accounts in large part for the prior sterility of these disciplines. Following the introduction of self-instructional test-scoring devices early in 1924, Pressey wrote: "The needful thing here is experimentally to compare learning 'by machine' with learning by more usual methods; a graduate student is now making this comparison" (p. 45). If studies of this type had been consigned exclusively to pre-doctoral research there would be less cause for concern. As Gilbert points out, however, there is currently "a whole rash of so-called 'control-experimental group' experiments purporting to answer questions about principles of programming education... [despite] a basis for more considered effort..." (p. 447). Several studies of the comparison-research type appear in "Teaching Machines and Programmed Learning." Porter has described the method and its limitations well:

"The procedure which has been followed is to obtain approximately equated groups of students and expose one group to the usual classroom methods of teaching...and the other group...to mechanical device teaching utilizing the same subject matter. Effectiveness of the two teaching methods is then evaluated by comparing the scores for the two groups of students obtained on identical tests.

"Such experimentation may indeed show an advantage for one or the other method of teaching, but there is no guarantee that the results obtained can be replicated, for the outcome of these experiments depends upon unspecified parameters of the 'usual' classroom situation. As stated by one group of researchers, 'the complexity of the teaching-learning process is such that attempts to establish the relative merit of a 'general method of teaching' are likely to prove inconclusive.' (Guetzkow _et al._, 1954). To be of value, investigations concerning mechanical teaching devices, or any other method of teaching, have to deal with the variables which lie behind the presumed superiority of the method" (p. 127).

(The author continues with a critique of the control-experimental group studies by Pressey and his co-workers.)

In the light of the obvious methodological limitations of comparison research it is difficult to understand what motivates its continued pursuit. The reviewer cannot agree with Carr that "a certain amount of evaluation research is necessary in order to justify continued interest in the basic concept of automated instruction" (p. 451). Interest in automated instruction is merely an extension of interest in the analysis and control of human behavior; it has the same justifications as the basic endeavor to understand man's condition and to improve it.

If further justification is needed, the reader may consider the likelihood that a systematic analysis of the acquisition of knowledge with the tools of a science of behavior will lead to improvements in current educational practices. For those who would "take the cash and let the credit go" there are cash prizes abundantly to be had, as the reports of "field trials" of programmed learning indicate. (See, for example, Blyth, p. 401.)

The basis for "a more considered effort" is the strategy of research that has led to a modern science of behavior. "The major portion of research effort should be devoted to an experimental analysis of the parameters which influence the effectiveness of self-instructional devices" (Carr, p. 541). Questions for research of the form: which is better, A or B? are not appropriate. Instead we should ask: under what conditions are A and B effective in controlling behavior? As Gilbert has said, we must ask, "What variable is effective? and what can teach?" (p. 484). In commenting on the proper length of programmed materials, Skinner has characterized this approach: "In the long run, only an experimental analysis of material in a natural class situation will determine suitable length for a given type of material" (p. 163). Several of the studies reported in "Teaching Machines and Programmed Learning" used this type of research approach: careful analysis of program and machine variables in terms of the "fine-grain" of student performance, followed by corrective adjustments in program techniques, content, and arrangement. Enough time has not passed since the inception of programmed learning for the products of such "iterative programming" to become widely available. The Holland-Skinner program, "A self-tutoring introduction to a science of behavior," may be the best example of iterative programming to date (vide p. 215ff).

The "more considered effort" in the improvement of educational practices referred to earlier should take place at two levels. Concurrent with an experimental analysis of variables influencing self-instruction, there must be continued research in the parent discipline: the experimental analysis of behavior. An analysis of behavior under the controlled conditions of the laboratory is propadeutic to the manipulation of that behavior in the complex environment of the classroom. (Vide Rothkopf, p. 328; Melton p. 663.)

REFERENCES

Guetzkow, H., Kelly, E. L. and McKeachie, W. J. "An experimental comparison of recitation, discussion, and tutorial methods in college teaching. J. Educ. Psychol., 45, 1954, 193-207.

Holland, J. G. and Skinner, B. F. The Analysis of Behavior. New York: McGraw-Hill, 1961.

Skinner, B. F. Verbal Behavior. New York: Appleton-Century-Crofts, 1957.

SOME DIFFERENCES BETWEEN FIRST AND SECOND LANGUAGE LEARNING

Harlan Lane
The University of Michigan

SOME DIFFERENCES BETWEEN FIRST AND SECOND LANGUAGE LEARNING[1]

Harlan Lane
The University of Michigan

Yesterday afternoon a young man interested in language learning, like each of us, came to me with a problem. He had read about my address this evening on some differences between first and second language learning, and he wanted to attend. On the other hand, he related he had promised to take a girl to the movies this evening. He didn't state the question bluntly but it clearly amounted to: which did I think would be more rewarding—my monologue or a movie.

I asked the young man if he had learned a second language in addition to English and he said he knew two foreign languages. "Why, then," I replied, "you can judge for yourself the differences between first and second language learning." "Yes," he answered, "I know some things from my own experience but you are a psychologist and could tell me much more." "Why so," I asked. He didn't answer, but with a puzzled look apologized for intruding and left. I don't know if he is here after all. Perhaps he struck a compromise between his two goals and took his date to a foreign film.

Our friend was clearly not a psychology major. A psychology major could readily answer my question: Why can a psychologist tell you more about language learning than you know from your own experience? An "A" answer on an examination might go something like this:

---

[1] An Address to the English Language Institute, The University of Michigan, April, 1960.

> The psychologist can tell you more about second language learning than you know from your own experience because a psychologist limits his experience. He studies limited samples of language learning under limited conditions. The relations between behavior and the environment are, therefore, simpler and more easily apprehended. A knowledge of these basic relations between language and the environment enables the psychologist to discriminate among relevant and irrelevant variables in the exceedingly complex language learning situation.

I would like this evening to put the student's answer to a test. How far will the findings of the laboratory and the concepts derived from these findings carry us toward an understanding of first and second language learning and their differences?

As soon as we attempt to characterize first language learning in terms of research findings we are at a standstill because of the first critical difference between first and second language learning. Second language learning is what we make it. First language learning is rarely planned or controlled. It is for this reason that psychologists and linguists have traditionally settled for a descriptive account of first language learning but insist on criticizing and improving upon second language learning. Although there is a dearth of studies concerned specifically with the infant learning to vocalize under controlled experimental conditions, our knowledge of the principles of learning based on research with other humans and subhumans behaving under controlled conditions may aid us in giving a plausible, if not proven, account of infant speech development.

Let us start with a description of early speech development in the child and then see what basic behavioral principles may be introduced to account for these developments. Since we did not participate in the manipulation of the child's speech we must inquire of the parent instead: What did you do to your child and what, in turn, did the little fellow do? Now here is a pretty mess. Most adults give very poor detailed accounts of their own behavior and distort extensively and variously in recounting the behavior of their children and the conditions which brought this behavior about. To quote from McCarthy in her classic review of the literature on language development in the child,

> "Although this wealth of observational material has proved stimulating and suggestive for later research workers, it has little scientific merit. For each of the studies employed a different method, the observations have, for the most part, been conducted on single children who were usually either precocious or markedly retarded in their language development, the records have been made under varying conditions, and most of the studies are subject to the unreliability of parents' reports."

A general outline of the development of speech in the infant may, nevertheless, be drawn from biographical accounts and from secondary sources such as those by McCarthy (1946) and Lewis (1951). Soon after birth, any stimulus produces a state of undifferentiated excitement in the infant. Many observers report that within the first few hours two "states" may be distinguished: distress and delight. To quote Lewis, "Each state is accompanied by a specific vocalization, crying in the former case and soft gurgling noises in the latter." Most writers agree that the differentiation of these affective states and associated reflexive vocalizing are the starting points in the development of speech.

The next major development in the vocalizing of the infant occurs some time during the second month of life when, among the sounds uttered in states of comfort, some babbling of isolated sounds appears. This babbling period continues for eight to ten months, during which time the phonetic structure of vocalizing is undergoing drastic but regular change (Irwin, 1941).

The third development that I shall single out in the acquisition of speech by the infant is called imitation. Although imitative behavior is usually reported after the ninth month, and seems to arrive abruptly on the developmental scene, Lewis suggests that its earlier traces may be observed concurrent with the development of babbling. It seems to be the consensus that the child imitates only those sounds that have already appeared in his babbling repertory; the imitation of the speech of others is then based on novel combinations of these sounds (Curti, 1938; Shirley, 1933; Guillaume, 1925).

Observations of subsequent linguistic development reveal an increasing complexity of performance which is equaled only by the complexity of the theories elaborated to account for it. Studies of language during the second year of life and beyond introduce such processes as the comprehension of speech, the mastery of conventional forms, the expansion of meaning, the development of reference to past and future, and so on. These topics take us beyond the present sketch.

We are, therefore, given these three highlights in the development of infant speech: (1) reflexive vocalizing, (2) development and articulation of the babbling repertory, and (3) imitation. This is, as you can see, a purely descriptive classification. Let us accept this synthesis of various descriptive sketches and see how plausible an account of these developments can be given in terms of behavioral principles.

Three basic principles must be introduced for our present discussion of first language learning. We will rely on these principles again in our discussion of second language learning. The first principle is reinforcement, the second is discrimination, the third is differentiation. The principle of reinforcement states simply this: a large part of human and subhuman behavior is controlled by its consequences in the environment. The consequences are called reinforcing events and the behavior which is controlled or changed is called operant behavior, since its defining feature is that it operates on the environment.

I am told that all great truths are immediately understandable. If the observation that behavior is controlled by its consequences seems eminently reasonable to you and hardly worth elevating to the rank of a principle, I invite you to consider how rarely we act on this understanding. Is the language learning situation engineered so that each student's behavior has immediate reinforcing consequences? Rarely so! And yet we would change the behavior of the student. And I, this evening, would like to change your behavior; loosely put, I would like to make you more aware of the underlying behavioral processes in language learning and more disposed to take advantage of this knowledge as learners and teachers. Yet do I permit you to operate on the environment? Clearly not. (The principle of reinforcement implies that I will accomplish more in the question and answer period than in the whole of this address.)

The second behavioral principle is discrimination. Behavior that is reinforced only under certain conditions will come to be emitted only under these conditions. This principle is readily demonstrated by the fact that one speaks French in French class, German in German class, and Jargon in psychology class. Or to use a more vivid example, one sings hymns in church and bawdy songs in fraternity houses and rarely the reverse—because of the reinforcing contingencies that obtain under these separate conditions.

The third principle we must introduce at this point, the principle of shaping or differentiation, provides that the form of a response may be altered by selective application of reinforcement, so that totally new responses may be shaped out of the current behavioral repertory.

Each of these three principles has been the subject of extensive laboratory research using humans and subhumans behaving under highly controlled conditions. Let us see now how much power these principles of operant control have in accounting for the three stages of infant speech development that I highlighted earlier: (1) changes in reflexive vocalizing or crying, (2) development and articulation of the babbling repertory, and (3) imitation.

The account is, of necessity, speculative. It is offered in the same spirit as the more comprehensive treatment of verbal behavior presented by B. F. Skinner (1957) and it would be well to quote his introductory remarks as a prelude here:

> "The emphasis is upon an orderly arrangement of well-known facts, in accordance with a formulation of behavior derived from an experimental analysis of a more rigorous sort. The present extension to verbal behavior is thus an exercise in interpretation rather than a quantitative extrapolation of rigorous experimental results."

It is in the selective reinforcement of crying that we find the first evidence of operant control of vocalizing. In a biographical sketch of his infant's speech development, Charles Darwin wrote: "After a time the crying sound differs as to the cause such as hunger or pain... he appeared to cry voluntarily." We see that crying is an early way of operating on the environment for the infant; the infant is reinforced for crying by the presentation of food or perhaps the removal of a wet diaper. This brief account of behavior also exemplifies the operation of discrimination and differentiation. Undifferentiated cries must have only a modicum of success. However, two responses of different form, each under discriminative control, that is—one cry when hungry, another when wet, have the effect of always producing the "right effect." As the parent learns to discriminate among the two cries he can more often respond appropriately. As a result, the differentiation of crying is reinforced.

If crying is reinforced frequently and intermittently it may pre-empt the development of other forms of social behavior in later months. Whining, prevalent in the older child, may represent a "regression" to an earlier form of successful vocal behavior. Williams (1959) reports the extinction of crying-at-bedtime of a child, 21 months old, by simply discontinuing parental attention to crying at this time. The extinction curves he presents resemble those for other human and subhuman operants.

In terms of the dichotomy proposed by Lewis (supra), I have suggested that the vocal behavior of the infant in a state of discomfort is amenable to operant control. It is unlikely, however, that crying is the raw material out of which complex speech is formed. A much more likely source for this performance is the babbling of the infant, associated with states of comfort. Irwin and Curry (1941) have recorded phonetically more than one thousand vowel-like sounds from forty babies observed during the first ten days of life. We have reason to believe, therefore, that sufficient variability exists in the very earliest repertory of the infant for the differential reinforcement of approximations to English.

Irwin and Chen (1946) have traced the number of native-tongue phonemes emitted by 95 infants in their home environments during the first three months of life. The mean number of phoneme types (arrived at by observer agreement) was found to grow as a negatively accelerated increasing function of the age in months. Although the mastery of phoneme types grows at a decreasing rate, the frequency of production of these phonemes is a positively accelerated function of age (Irwin, 1947). Most biographical accounts concur with the more rigorous empirical studies performed by Irwin and his colleagues in reporting an overall increase in the frequency of babbling and increasing approximation of the babbling repertory to English (McCarthy, 1946; Lewis, 1936; Leopold, 1939).

If we were to attribute the former finding, the increase in the rate of babbling, to operant control, it would not be entirely speculative. First, we have an analogous finding in experiments with chicks, parakeets, and cats; we know that the rate of subhuman "babbling" may be manipulated by reinforcement (Lane, 1961; Ginsburg, 1960). Furthermore, Rheingold, Gewirtz, and Nelson (1959) have demonstrated the operant conditioning of babbling in 21 infants, median age, three months. Regular reinforcement (smile plus three "tsk" sounds plus a light touch applied to the abdomen) of vocalizing produced an increase of over 100 per cent in the number of vocal responses per session, while discontinuing reinforcement led to a drop in responding back to the original baseline level.

In order to account for the increasing articulation of the babbling repertory, however, we must introduce the notion of selective reinforcement: We assume here that the child's verbal community is under the discriminative control of the child's speech with respect to its reinforcing practices. A mere disposition to reinforce the child for vocalizing at all is not sufficient. We are assuming that planned and unplanned contingencies operate selectively to enhance the strength of English approximates and to neglect or extinguish non-English sounds. When the child speaks English, we act and his speech has a reinforcing effect. When he speaks nonsense we call it senseless and rarely reinforce.

Selective reinforcement of responses appearing in the babbling repertory may be responsible in large part for the increasing approximation of the infant's phoneme repertory to that of the adult, linguistic community. Furthermore, relatively simple words and compounds in the two-year-old's vocabulary are probably differentiated directly out of the babbling repertory. Since babbling is characterized by short, repetitive sequences, we may expect reduplicated monosyllables, such as ma-ma and pa-pa, to arise earliest directly from this repertory, and without imitation. Baker (1955) is lead to related conclusions from an etymological analysis:

> "This interlocked issue of appropriations by elders and the weight of conditioning imposed by the linguistic community into which the child is born, operating as they do to shape spontaneous infant vocalizations into phonemic forms, is highly complex both in its range and products. We have seen how, in certain words for father, p and b sounds have been interchanged. Precisely the same thing happens with t and d sounds, both of which (once again) Lewis has recorded among infant utterances. Compare English dad, Welsh tad, Irish daid, Breton tat and tad, Greek tata, Sanskrit tata, all applied to father. And from the other side of the

> world: Sentani *adai*; Malagasy *dada* and *daday*; Fiji *ta* and *tata*; Pampang and Guaham also have *tat* for father; in Formosa *ta* is used as a prefix for the names of men.
>
> "What is being suggested here is that infant vocalizations—the spontaneous and instinctual utterances that the child brings into the world—form the matrix of language. (Not all words), but certain nuclear words are formed by and drawn from the matrix of infant utterances" (p. 328).

Once a basic repertory begins to develop, vocal behavior will tend to be reinforced in preference to other motor behavior:

> "At the same time that the child is being rewarded for making more responses to words as cues, he is gradually learning another aspect of language, namely, how to make the response of uttering words. If a cooky is out of reach the response pattern of pointing at it with the body and eyes and reaching for it with the hand is often rewarded by inducing some older person to give the child the cooky. If this gesture is accompanied by a sound, it is more likely to be rewarded. If the sound seems to be some appropriate word, such as 'Look at,' reward is still more likely. Eventually the more effortful parts of the gesture drop out, and the verbal response, which is least effortful and most consistently rewarded, becomes anticipatory and persists. The mechanism of reward gradually differentiates language from it original matrix of other, more clumsy, overt responses. The child learns to talk because society makes that relatively effortless response supremely worthwhile." (Miller and Dollard, 1947, p. 82)

You may agree at this point that our principles of operant control account well for the development of the elements of speech in the infant. But how to deal with the more advanced process of imitation? Imitation is generally given the lion's share in an account of the development of speech and is the third major development in the acquisition of speech by the infant that we noted earlier. One use of the word as an explanatory concept is clearly circular, and this facile circularity has no doubt contributed in large measure to the popularity of the term. The datum to be accounted for is the increasing complexity of the child's speech or, in other words, the increasing approximation of the child's speech to that of his elders. Descriptively, the child comes to *imitate* the vocal behavior of the linguistic community and especially that subcommunity which his parents comprise. An explanation of this imitative behavior by reference to the process itself gives the circular account: a child imitates because he imitates.

Lewis (1936) describes the development of imitation in this way:

> "...for a very long time the forms used by the child in imitation of adult language consist of his own familiar sounds spoken as approximations to those that he hears. Only gradually, as he attends more closely, are the movements of his vocal organs subordinated to his auditory perceptions. At first he is satisfied to make broad, crude attempts: as time passes his vocal movements become more and more refined. Slowly he comes to pronounce his mother tongue in the accepted fashion, under the stress of social selection, *that is, the responses made to his attempts by others*" (italics mine).

Lewis' description exemplifies what we have called differential reinforcement of verbal behavior. Once again, we may point out that the positive disposition of the parents to reinforce "proper speech" facilitates this acquisition process, for it is primarily the parents who respond to the child's vocal attempts. Increasingly accurate approximations by the infant to the language of the community are reinforced not only because they are likely to be more effective (more rapid, more reliable) in parental control, but also because parents often actively shape the speech of their progeny at this stage of linguistic development.

As B. F. Skinner has put it:

> "Echoic behavior, like all verbal behavior, is shaped and maintained by certain contingencies of reinforcement. The formal similarity between stimulus and response is part of these contingencies and can be explained only by pointing to the significance of the similarity to the reinforcing community." (1957, p. 59)

This fact is rather entertainingly underscored in a passage from Samuel

Butler's Way of All Flesh:

"Ernest," said Theobald..., "don't you think it would be very nice if you were to say 'come' like other people, instead of 'tum'?"

"I do say tum," replied Ernest...

Theobald noticed the fact that he was being contradicted in a moment...

"No, Ernest, you don't," he said, "you say nothing of the kind, you say 'tum', not 'come'. Now say 'come' after me, as I do."

"Tum," said Ernest...

"...now, Ernest, I will give you one more chance, and if you can't say 'come' I shall know that you are self-willed and naughty." (cited in Skinner, 1957, p. 60).

To summarize, our account of infant speech acquisition in terms of reinforcement theory develops along the following lines:

1. Crying and babbling occur at a high unconditioned rate in the earliest hours of an infant's life.

2. There is some selective reinforcement of crying, so that it presently comes to function as a mand and to exert social control.

3. There is generalized reinforcement of babbling so that it increases in rate during the first year.

4. There is selective reinforcement of babbling so that the phonetic structure of the babbling repertory comes to approximate that of the language. Furthermore, certain elemental words tend to occur as a result, are reinforced, and increase in frequency.

5. Adults generate a great deal of vocal behavior in the presence of the

babbling child. In accordance with step 4, there is considerable overlap between the phonetic structure of the child's vocalizing and that of the adult. When a babbling response is emitted that has some formal similarity to the vocal productions of the adult, it tends to be reinforced.

6. As a result, phones emitted by the adult tend to evoke similar phones emitted by the child. Novel words emitted by the adult tend to evoke their phonetic components.

7. Approximations to the words of adults emitted by the child are reinforced. As the vocabulary of the child increases in breadth, the criteria for a "good approximation" and hence the contingencies of reinforcement become more stringent.

If the principles of operant control are at work in first language learning it is clear that they are not employed to full advantage. As parents we are inconsistent in our reinforcing practices. We permit correct responses to go unreinforced and fail to reinforce desired behavior. Furthermore, reinforcement practices are inconsistent from home to school and from school to street in later stages of speech development. That we have some success, nevertheless, is testified to by the many Americansthat speak English. That we are grossly inefficient is testified to by the differences in verbal prowess among individuals and across socio-economic levels.

Practically speaking, we need not engage in these undesirable practices in teaching the second language; once again, this is the overriding difference in the learning of these two languages. We can and we will take advantage of scientific knowledge in arranging second language learning.

A second difference between first and second language learning is in the nature of reinforcement control. In second language learning we must rely on such spurious reinforcers as a nod, a smile, a little approval. Most of all—it must be admitted—we rely on punishment and the threat of punishment. The grade and the prerequisite serve us as well—or as poorly—and little more subtly than the birch rod served our forebears. Our reliance on punishment is an explicit acknowledgement of this difference between first and second language learning. We do not have the absolute control of the parent over the child, nor the use of primary reinforcers such as food, and we fear or find that secondary reinforcers such as approval will not serve alone.

A third difference derives from the fact that the student learning a second language begins with a highly articulate verbal repertory. This verbal ability is usually seen as expediting the second language learning process but in particular cases the two repertories may actually conflict. The clearest example of repertories in conflict occurs when the second-language learner is confronted with a foreign word that has an English cognate or that has been "borrowed" into the English language. Language programmers tell me that they leave such words as "mesa" and "adios" in Spanish, and "bonjour" and "parlez-vous" in French for very late stages of their programs when vocal skills are well mastered, and the tendency to say these responses as an American is relatively weak compared with the tendency to render the correct pronunciation. Similarly, many language teachers report that the introduction of "realia," or "meaning," or Latin orthography, usually leads to a decrement in pronunciation. We may expect that this degradation is due to the elicitation of English vocal responses by these stimuli, whether objects, concepts, or letters. These English responses then compete with, or even override, the newly formed foreign responses with the result that pronunciation is impaired.

The fourth and final difference between first and second language learning that I should like to point to this evening, I believe to be the most critical and the least widely known. The nature of this difference has become clear to me only after some six months of research in conjunction with the Language Laboratory here at The University of Michigan. This critical difference is in the nature of discrimination learning. Earlier in this address, I stressed the importance of discrimination learning in the development of the first language. It is the process by which one learns to say the right thing at the right time. Imitation is dependent upon discrimination, as are most vocal skills.

The process by which behavior comes under stimulus control initially is a gradual one. Now it is difficult if not impossible to study initial discimination learning in humans, for this requires a naive organism, to use the technical sense of the word. There seem to be three courses open to the researcher: first, he can employ very young infants; however, in addition to the obvious ethical problems impeding research there is the fact that the child very early comes to discriminate the components of the "blooming, buzzing confusion" that confronts him upon entering the world. Second, the behavioral scientist can employ adults, and attempt to study discrimination learning under conditions where prior discrimination learning is not relevant. This has probably never been done, since the adult has an extensive and variegated history of discrimination learning. Finally, the researcher can employ subhumans, whose training history he can control. This approach to understanding discrimination learning has been pursued extensively, and the finding is, as I have said, that initial discrimination learning proceeds slowly.

Allow me to describe the course of discrimination learning of vocal behavior in the chicken and then to contrast this initial discrimination learning with the analogous process in second-language learning. At first, we bring the vocal response of the chicken under reinforcement control. We may increase or decrease the rate of chirping at will by appropriate contingencies of reinforcement. Then, to bring the response under discriminative control, we set up reinforcement contingencies that are unique to the stimulus conditions. For example, when the word "chirp" is played repetitively to the chicken we reinforce chirps, by presenting food to a food-deprived chick contingent upon chirping. When the words "do not chirp" are presented, chirps have no consequences in the environment, they are not reinforced, chirping is, so to speak, extinguished. Now, observe the course of discrimination learning. Gradually, chirping in the no-reinforcement condition extinguishes. Over the course of a few hours, the rate of chirping in this condition may fall to zero. In the chirp condition, however, where responses are reinforced, the rate remains quite high. Thus by the end of the experiment, the bird chirps when the chirp stimulus is on and rarely or never chirps when the "do not chirp" stimulus is in effect.

Now let us examine the analogous experiment in auditory discrimination learning with second-language learners. For example, we present a Spanish phone, such as /a/; if the subject responds to this Spanish stimulus by saying

"Spanish" or by pressing a button, he is reinforced—with points or the bleep of a tone. Then, too, there are negative stimuli, when responding is not reinforced. These are English approximate sounds such as /ae/. Here, too, the subject learns to discriminate one auditory stimulus from another. But now, the big difference: the process is not gradual. What we observe instead is a few trials on which errors occur and then, abruptly, the student is one hundred per cent correct. He always responds to Spanish and never to non-Spanish. Why the big difference? Why isn't discrimination learning in the second language gradual? The answer is: because the student has already learned to make these discriminations in the course of learning his first language. He can "tell the difference" between /a/ and /ae/ just as you can. Indeed, he can tell the difference between allophones of the same phoneme, by virtue of his prior training. As a result, the errors that the student makes in second-language discrimination learning are usually errors of over-discriminating. He fails to respond to variants of the positive stimulus which the experimenter considers equivalent.

Mr. Dale Brethower has recently demonstrated this nicely with a non-Latin language—Thai. Students were given the task of simply saying whether two sounds were the same or different. The sounds of the pair were either both Thai, or one Thai sound and one English approximate. The finding: most Thai sounds, even the most difficult, have proven to be discriminable. There were no Thai sounds that all subjects failed to discriminate from their English approximates. You see, in learning to discriminate among the sounds of a second language, the subject is not learning a discrimination at all. He is learning to transfer discriminations that he is already capable of. As soon as he knows your set of rules, so to speak, he plays the game perfectly.

This phenomenon is not new to the psychological literature. Whenever a subject is given the task of learning a discrimination for which he has extensive prior training, the learning process is abrupt. For example, in an experiment by Heidbreder (1947), subjects had to learn the nonsense syllable names of a group of objects and abstract forms. They were already quite capable of discriminating among the objects and forms, such as faces, animals, colors, and so on. What they did not know was that certain of the obvious distinctions among these stimuli were irrelevant. To be right, it was necessary to consider a variety of animals, for example, as equivalent, and give the same nonsense syllable name to each. The subjects' errors were, as in the case of second-language learning, errors of over-discrimination. The subject was capable of discriminating among allocons of the same concept, so to speak, although by definition, these differences were irrelevant. As a result, the learning curve shows many errors for a short while, and then an abrupt increment to perfect performance. The time from the first correct guess to one hundred per cent correct naming was usually one or two trials. Contrast this with the thousands upon thousands of responses that are required in initial discrimination learning, before the discrimination is mastered. Heidbreder calls the process of transfer of earlier discriminative behaviors "concept attainment."

I believe that an appreciation of these differences between first and second language learning that I have singled out this evening should color our techniques as second-language teachers to a large extent. Allow me to recapitulate these differences. First, there is a great difference, practically speaking, in the measure of control that we can exert over first and second-language learning. Second, there is a great difference in the nature of the reinforcers that are available to us. Third, we must remember that the second-language learner, unlike the infant, has a highly articulate verbal repertory. Fourth, we must remember that the second-language learner, unlike the infant, has had extensive discrimination training and is essentially faced with the task of "concept attainment" rather than discrimination learning in coming to respond appropriately to the sounds of another language.

May I repeat that these differences should color our technique as second-language teachers. I would be very pleased if the effect of my lecture this evening were twofold: first, the development of a greater awareness of the basic behavioral principles that can be employed to optimize second-language learning. In particular, the principles of reinforcement, discrimination and differentiation. And second, a greater awareness of the student's point of departure in second-language learning: his discriminative abilities and his current vocal repertory.

REFERENCES

Baker, S. J. The instinctual origin of language. *J. gen. Psychol.*, 1955, 53, 305-328.

Ginsberg, N. Conditioned vocalizing in the Budgerigar. *J. comp. physiol. Psychol.*, 1960, 53, 183-186.

Heidbreder, E. The attainment of concepts: III the process. *J. gen. Psychol.*, 1947, 24, 93-138.

Irwin, O. C. Research on Speech sounds for the first six months of life. *Psychol. Bull.*, 1941, 38, 277-285.

Irwin, O. C. Development of speech during infancy: curve of phonemic frequencies. *J. exp. Psychol.*, 1947, 37, 187-193.

Irwin, O. C., and Chen, H. P. Development of speech during infancy: curve of phonemic types. *J. exp. Psychol.*, 1946, 36, 431-436.

Irwin, O. C., and Curry, T. Vowel elements in the vocalizations of infants under ten days of age. *Child Developm.*, 1941, 12, 99-109.

Lane, H. L. Operant control of vocalizing in the chicken. *J. exp. anal. Behav.*, 1961, 4, 171-177.

Lewis, M. M. *Infant speech*. New York: Humanities Press, 1951

McCarthy, D. Language development in children. In Carmichael, L. (Ed.) *Manual of child psychology*. New York: Wiley, 1946.

Miller, N., and Dollard J. *Social learning and imitation*. New Haven: Yale University Press, 1947.

Skinner, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

Williams, C. D. The elimination of tantrum behavior by extinction procedures. *J. abnor. soc. Psychol.*, 1959, 59, 269.

TECHNIQUES OF OPERANT CONDITIONING APPLIED TO SECOND LANGUAGE LEARNING

F. R. Morton and H. L. Lane
The University of Michigan

TECHNIQUES OF OPERANT CONDITIONING APPLIED TO SECOND LANGUAGE LEARNING[1]

F. R. Morton and H. L. Lane
The University of Michigan

Mr. Rigney's summary of programs and program analysis is helpful to the writers of the present paper because it emphasizes the great distance that separates conventional concepts and research in teaching machines and programmed learning from those that we are about to report. In preparing their summary of automated, self-instruction programs in the United States, Mr. Rigney and his associates were concerned primarily with the means for shaping covert verbal behavior. The overt correlates of this behavior, required by the teaching device itself, have been shown by many investigators to be entirely contingent upon changes in covert verbal behavior. The programs that Mr. Rigney has categorized all have in common that, speaking literally, they do not involve conditioning at all. In terms of a change in behavior, there is either none or only the most superficial kind, that of changing a general vocabulary to a specific or technical one. This type of verbal conditioning, involving as it does a mere restructuring of the subject's extant verbal repertory, may be contrasted with the type of conditioning we have undertaken in our programming of audio-lingual behavior. Programming for the acquisition of second language fluency requires verbal conditioning in the strict—not extrapolated—use of the word. New auditory discriminations must be conditioned, new patterns of vocal behavior must be differentiated, and discrimination and differentiation must be coordinated to provide the conditioning of the complex skill that is the desired terminal behavior. The conditioning tasks that compose our audio-lingual program are indeed indistinguishable from those undertaken in the operant conditioning laboratory. Since our applied techniques are based on the principles developed in the laboratory, the similarity of the tasks augurs well for the applicability of the principles. This notion, that audio-lingual programming represents not an extrapolation but a generic extension of operant conditioning techniques has been amply verified by our findings.

---

[1] An address to the International Congress of Applied Psychology. Copenhagen, August 17, 1961.

Two programs of research have been pursued and are currently in progress. On the one hand, the techniques of operant conditioning have been applied to second-language learning in a heuristic problem: conditioning a rat to discriminate among spoken languages. The traditional operant discrimination learning paradigm was employed. During the positive discriminative stimulus, a 30-second English passage, every tenth bar press provided the rat with a little sweetened condensed milk. During the negative stimulus, a comparable Spanish passage, each bar press shut off the apparatus for ten seconds, thus postponing the occasion on which the rat could earn more milk. The English and Spanish passages were of the same over-all intensity, duration, and pitch and were presented in random order. As you no doubt anticipate, the rat soon learned to respond only when English was presented, that is during S^D and never when Spanish was presented, that is during S^Δ.

The terminal behavior of our rat impressed many onlookers but puzzled us. What components of the complex stimulus patterns were controlling his behavior? This was the first question that came up and it returned again when we set out to program human discrimination of foreign language sounds. A related question that arose was: what were the sources of generalization between the two patterns that retarded the development of differential responding? This question was to be raised again in teaching members of the English-speaking community to discriminate among the sounds of English and Spanish. Many observers claimed that our rat obviously understood the language passages presented. Our first tendency, like yours, was to say certainly not, but further consideration suggested otherwise. Our rat was indeed responding appropriately, the earmark of understanding; his behavior was sandwiched between the discriminative stimulus and the reinforcing stimulus in a highly predictable way. This three-term relation has been identified by B.F. Skinner as the foundation of verbal behavior. In any event, a third central question was now before us: just what behaviors do we require before we say that a student understands a foreign language?

The contingencies of reinforcement that were employed in training our rat were selected so as to minimize responding in S^{Δ} and provide a high rate in S^{D}. Other contingencies would have yielded a more rapid but less stable development of differential responding. Clearly, a fourth question was raised: having specified the terminal behaviors desired (and the constraints of equipment and time) what contingencies of reinforcement are optimal?

With at least these four questions in mind we undertook to program the acquisition of second-language fluency. First, the repertory of linguistic behaviors in the American student was summarily noted. Second, the terminal behaviors desired at the end of the conditioning process were laid down in the fullest detail; the auditory discriminations and vocal productions required were enumerated, based on the findings of structural and descriptive linguistics. Finally, a program was prepared, leading from the extant to the desired repertory by small steps, in increasing order of difficulty and complexity.

The terminal linquistic performance may be artificially, but conveniently, categorized into four sub-repertories which are conditioned in this order:

1. Acoustic (phonetic) discriminations. Here, of course, our primary purpose is to enable the subject to hear "correctly" the new sounds of the foreign language, to discriminate between them and the sounds of his own or native language. We have experimented with numerous techniques to accomplish this task and shall describe but one. The subject hears a group of five or six speech sounds of the target language mixed with approximate but non-target sounds. His task is to respond to target sounds, not respond to non-target ones. The motor response employed in this case was the pulling of a manipulandum. If the subject responds correctly by pulling to an S^D, or failing to pull to an S^Δ, he hears a confirmation tone and receives a point on a counter. If he responds incorrectly, he hears no tone and looses a point. An arbitrary criterion of accuracy is required before the subject can proceed to a new set of target and non-target sounds. He achieves criterion in a

remarkably short time. In one case, three students were conditioned to discriminate 28 Spanish phonemes from some 62 non-Spanish approximate phonemes in less than 8 hours.

2. Acquisition of vocal responses. Once the student has been trained (i.e., conditioned) to discriminate the new, target sounds of a language, this behavior is utilized by the program to shape vocalization of the same sounds. Again, only one of the techniques experimented with can be mentioned. Here the subject hears the model sound in his earphones; replicates it as best he can then hears immediately played back to him the original model, his attempt at its replication and again the original model. Shaping of the student's echoic response to criterion accuracy requires between 60 and 180 echoic responses for all subjects thus far run, some 90 responses on an average (a period of about five minutes). Elicitation of the vocal response is then conditioned to secondary auditory and non-auditory stimuli until the student is able to generate the sound or sounds in question under a variety of circumstances while maintaining his original skill of discrimination between target and non-target sounds. At the conclusion of this task our subjects have been able to replicate any short (up to 12 syllable) utterance in Spanish with high phonetic and prosodic accuracy.

3. Syntactic or structural discrimination and production. This third task returns to the techniques employed in the first with the exception that the discriminative response to a structural or syntactical S^D is a pre-determined vocal response on the part of the subject. In this way is taught the so-called "acoustic grammar" of the language in question. At the conclusion

of this task our subjects are able to respond correctly (that is, to behave as a native might behave) to many verbal stimuli presented to them, although lexical meaning has not been introduced.

4. Model pattern performance. In this last stage the student learns to integrate the first three performances. When asked a question, for example, in the target language (requiring phonetic and structural discrimination), the student responds in that language (requiring differentiation), in an acceptable and meaningful form (requiring the coordination of these skills). It is at this stage that lexical "meaning" is introduced.

A series of subjects who have undergone this program show rapid mastery of second-language fluency. Small scale experiments in various stages of the program reveal the extensive control over component behaviors exerted by the program and point the way toward improvements based on the disparity between obtained and ideal terminal behaviors.

EXPERIMENTATION IN THE LANGUAGE CLASSROOM:

GUIDELINES AND SUGGESTED PROCEDURES FOR THE CLASSROOM TEACHER

Harlan Lane
The University of Michigan

EXPERIMENTATION IN THE LANGUAGE CLASSROOM: GUIDELINES AND SUGGESTED PROCEDURES FOR THE CLASSROOM TEACHER

Harlan Lane
The University of Michigan

Behavioral Technology and Language Learning

A technology of behavior is no longer a dream, but a reality. As a result of the growth of behavioral science, human behavior is now being "engineered" in the classroom. A language teacher may realistically envision the day when the student considers language learning an opportunity rather than a dismal fate. The "language block" will no longer mean an inability to learn, or a place of execution, but rather a core group of languages and language skills that the student readily and eagerly masters.

Learning: Sunburn or Behavioral Change

Scientists are beginning to develop an image of the optimal learning situation. The teacher may not be surprised to discover that current and traditional pedagogical techniques are greatly at variance with this image. Many teachers continue to be burdened with the "sunburn" model of learning. The teacher, prime source of knowledge, light (and, occasionally, heat) "exposes" students to his ideas; they "soak it up" and, in turn, become "enlightened." Students who fail to learn are simply not "sensitive" or "receptive," they do not "see the light." A newer, more workable model is emerging from current behavioral research. This characterization defines learning in terms of a change of behavior.

Consider the student who is about to learn French. He does not distinguish properly among French sounds; he does not respond appropriately when addressed in French; he does not produce most French sounds correctly; he cannot read French from a text; and so on. The teacher's task is to modify the student's behavior so that he will hear, understand, speak, and read French. To change the student's behavior from what it is now to what it should be: (1) the student's current behavior must be carefully assessed; (2) the desired terminal behavior must be carefully analyzed; and (3) a program must be set down that will lead in small steps from initial to terminal behavior. A characterization of learning in terms of behavioral change further requires that both student and teacher actively and profitably engage in the learning process. The student must respond if his behavior is to be changed, and the teacher must be alert always to insure that the behavior has some positive consequence, some effect. The teacher is clearly in the business of controlling the student's behavior: accepting each step forward, rejecting each step backward, he shapes the current behavior of the student gradually until it comes to approximate the terminal behavior.

Exactly what are the desired terminal behaviors in language learning? Descriptive and structural linguistics are providing an account of the terminal behaviors that are required for foreign language fluency. How can these terminal behaviors best be developed from the initial repertory of the student? Psychology is building the bridge between initial and terminal behavior by specifying programming techniques that will facilitate learning.

What role can the language teacher play? The language teacher can conduct important research within the context of the traditional classroom. Lest we too quickly abandon tried (if not true) methods and succumb to nothing more than a fad, we must use the classroom as a proving ground for new techniques. Furthermore, small-scale but rigorous research in the classroom can generate a wealth of provocative ideas and experimental findings.

A most valuable resource in improving modern language pedagogy is, therefore, you—the language teacher. This article has been written with the hope of stimulating your interest in the techniques and findings of behavioral science and in the pursuit of research in the language classroom.

Some Questions for Classroom Research

Each class hour can be part of a learning experiment. You introduce a controlled change in technique or content and observe a related change in the performance of your students. No matter what the outcome of this experiment, if you know what you did and what your students did, you can make some positive statement. In this sense a properly performed experiment always "works."

There are no absolute rules for generating good experiments, but a recurrent feature is that the experimenter is interested in the experiment; he is curious about a question that the experiment will answer. Perhaps some of the following questions will seem interesting to you; worthwhile asking and answering, and will prove suggestive of other experimental questions.

1. What would happen if...your Russian students learned Cyrillic script from a specially prepared program? When you are ready to teach orthography

in your course, you section the class at random into three homework groups. Group A, the control, is assigned the task of copying the dialogue appearing in Cyrillic in the textbook; they are to do "the best they can" and to hand in their work the next day. (This may be the technique you are using now.) Group B learns Cyrillic script from a "program" that you specially prepare. Here's how you might do it: Bear in mind the writing skills that the student now possesses and those that you wish to develop (the "terminal behavior"). Based on your experience as a teacher write out a sequence of symbols in increasing order of difficulty. The first symbols may not be Cyrillic "letters" at all, but parts-of-letters that are not difficult to draw. Do not be afraid of too slowly increasing the difficulty of the symbols you choose. (Almost every programmer begins by increasing the difficulty of his teaching program too rapidly.) After this sequence of parts-of-letters and letters is completed, join the letters into groups of two's and three's, then into words and, finally, sentences. This is your "program" for teaching Cyrillic script. To arrange that the students' behavior have some consequence at each step you might try this: Write all the symbols in order on index cards (and number them). Leave every other card blank. The student is instructed to examine the stimulus card, turn it over, write his response on the next (empty) card, and then compare the two. Then, he is to go on to the next stimulus card and proceed in this manner through the pack. On the following day, the student turns in his work so that it may be graded. Your third experimental group (C) can do both: work through the script and copy the text.

Your "independent variable" was the script program. What shall be the dependent variable—what change in behavior should you measure? Perhaps someone, unacquainted with the experiment, will grade the work of the three groups for you and you will compare their average grades. You may also use other measures of learning. For example: By administering a writing test at a later date, you can determine how the three groups compare in their ability to retain the writing skills they have mastered.

The time required to do the homework should be roughly equal for the three groups. If you "control for" this variable, it will not confound your results. This is an example of exercising experimental control. It is reasonable to assume that time spent in learning script, by whatever method, affects performance on a writing test. Let us say that your three groups learned script by the different methods and also spent different amounts of time in learning. Suppose, that the group scores on the writing test were found to be different. Are these differences in score due to different learning methods or to the different amounts of time spent in learning? As you can see, the effects of these two variables—method and time—would be confounded in your results.

What would happen, after all, if your Russian students learned Cyrillic script from a specially prepared program?

2. What would happen if...one of your Spanish classes learned the first three or four beginning dialogues from a text that had numbers in place of vowels? Since English and Spanish use similar written symbols, you may have observed students who use English sounds in response to the letters in their

Spanish textbook. One way of preventing this transfer of English speech habits in the reading of Spanish text is to remove the stimuli that elicit the English responses—namely, the letters common to both languages. You might want to try a competely new, arbitrary symbol system. Short of this, the present experiment proposes that you try removing the most common symbols and source of trouble, the vowels. Copy the first few dialogues in the text onto a mimeograph stencil, substituting "1" in each place that "e" occurs, "2" for "u," and "3" for "a," and so on. As you have done perhaps in prior courses, read the Spanish materials aloud (you may need the original text for this) and drill your students in pronunciation. If you have a second class using the unaltered textbook, these students may serve as a control group. The details of the experimental design and the choice of a dependent variable are left to you.

3. What would happen if...you taught French vocabulary with pictures? One group of subjects learns French words in response to pictures only; a second group learns French words in response to their English "equivalents;" a third group is presented with both the pictures and the English words when learning French vocabulary. How would these groups compare on a subsequent vocabulary test? How would they compare on a retest several weeks later? (Or, better, how would they compare if the first test were postponed a few weeks?) And, incidentally, how would the experimental and control groups compare on a test of pronunciation ability for these words?

4. What would happen if...you used the SRS model in preparing your language laboratory tapes? In line with our earlier distinction between two

conceptions of learning: sunburn vs. behavioral change, you may now be merely "exposing" your students to a second language in the language laboratory. What would happen if your tape recordings were prepared in this manner: first, the acoustic stimulus (S) (an isolated sound, a word, phrase, or sentence), then a pause during which the student gives an imitative response (R), then a repetition of the stimulus (S), yielding "confirmation." Again, the details of design and choice of a dependent variable are left to your ingenuity.

5. What would happen if...(for administrators only) your language teachers were given easier access to the professional literature in their field? Select a few important journals and enter several subscriptions. Distribute these personal copies to half your teachers (you may want to give them a copy of this article as well). Do nothing to the other half of your teachers, allow them to continue in their normal reading practices. At the end of the semester, distribute a rating form to the students (and/or their parents) in all classes. Ask them to rate the teacher on such dimensions as versatility, initiative, ingenuity, enthusiasm, and so on. Then, compare ratings.

6. What would happen if...

(Left blank to be filled in by the reader)

On the Significance of Results

Since you are actively engaged in language teaching, you probably have an image of the ideal language-learning situation and you may be convinced that it exists rarely, if at all, in our classrooms. (As indicated earlier, the psychologist will readily agree.) If you are willing to allow that there is great room for improvement in language teaching, you will probably agree that the only important changes in technique are those that show dramatic effects. At this point in our knowledge, changes in the learning situation that produce marginal changes in behavior are not significant—in the sense that they are not very interesting. These "small effects" may, however, encourage you to further research along the same lines. Small effects often grow to become large ones when the experimenter "refines" his technique and extends his control to more of the learning situation.

In addition to the "size" of an effect, there are other criteria you may take into account in estimating the importance of your findings. "Reasonableness" is one. Do the results of your experiment "make sense"? Do they agree with other experimental findings? If they do not, you may be on the verge of a new discovery and will want to check up on it with further research. More likely, however, you have made an old discovery—some uncontrolled variable is wreaking havoc. As an example, consider the experiment on programmed learning of Cyrillic script. You will remember that Group A copied the text, Group B received the script program, and Group C did both. Suppose that, in the writing tests, Group B did the best, Group A second-best, and Group C poorest. These results don't quite "make sense;" you may wonder how to account for them. If programmed learning (Group B) is better than copying (Group A), why should both combined (Group C) give poorest performance? One possibility is that Groups A, B, and C were not truly comparable before the beginning of the experiment and their penmanship grades reflect two confounded variables: learning method and prior skill.

You will observe that the criteria for importance of results, how dramatic are they and how reasonable, draw heavily on your experience as a language teacher and on your knowledge of psychology and linguistics. There is no other course; it takes experience and knowledge—that is, sophistication in your field—to assess properly the importance of your findings.

The size of an effect and its reasonableness tell you something about its *reliabiliy*, too. A reasonable but small effect will probably turn up again in the same or similar experiments. A reasonable and large effect is even more likely to recur. If another person who does not share yourprivate sophistication wishes to assess for himself the reliability of your findings, he has two courses open to him. First, he may replicate your experiment and see if he gets the same results. Alternatively, he may use a public criterion of reliability, employing statistics. Many psychologists publish statistical tests of their findings along with their report of research with this reason in mind—to aid the uninformed reader in arriving at an opinion about the reliability of their findings. Essentially, the statistical tests (unfortunately called significance tests) tell you what the odds are that the difference between your experimental and control groups is just a chance happening.

There are many pitfalls in applying statistics in assessing the significance of data. Perhaps, the most dangerous is that your devotion to statistics, may deflect interest from the practical and theoretical importance of your findings, which are quite another matter. Statistical significance does not guarantee either practical or theoretical importance. You incur a second danger in selecting a statistical test to be applied; often, statistics are applied to data for which they are not suited. Undoubtedly, the best course to follow, where possible, is to increasingly refine your technique and experimental control until your results are unequivocal.

Telling the World

There are many advantages in informing others of your experimental findings. In addition to receiving prestige as a researcher, you may receive helpful criticisms of your experimental design, references to related studies by other experimenters, indications of the range of applicability of your findings, and suggestions for follow-up research. In preparing your findings for publication, you may want to include the following steps: (1) tell others about your work; (2) then, write it up informally and distribute a dittoed copy to your fellow teachers and to someone who is trained in research methods, such as a psychologist or a linguist; (3) look over the journals in your field, and consider which one contains articles like your own; which one is read by the audience you wish to attract. Note the format in which the articles are presented and bear this in mind in your "write-up." Most journals have a manual of style to which you can refer. (4) Submit your article to the journal! Remember that the editors can also aid you in preparing the final manuscript by criticizing both form and content. Since it is true that neither piety nor wit will serve to retract an article once published, we strive for perfection before publishing. Nevertheless, suggestive findings from small-scale experiments deserve communication as well as the more definitive findings from large-scale research.

Some References

1. Carroll, J. D., Research in teaching foreign languages, *Handbook of Research in Teaching*. (Preprint, 1960).
2. Lumsdaine, A. A. and Glaser, R., *Teaching Machines and Programmed Learning: a Sourcebook*. Washington, Department of Audiovisual Instruction, National Education Association, 1960.
3. Morton, F. R., *The Language Laboratory as a Teaching Machine*. Ann Arbor, The University of Michigan Language Laboratory, 1960.
4. Skinner, B. F., The science of learning and the art of teaching, *Harvard Education Review*, 24, 1954, 96-97.
5. Skinner, B. F., Teaching machines, *Science*, 128, 1958, 969-977.

RESEARCH IN PROGRESS

Research in Progress

1. Foreign accent and speech distortion.

Four foreign students, with minimal training in the English language, read phonetically balanced lists which were presented to American undergraduates for intelligibility testing. Several types of distortion were introduced during the presentation of the lists and the effects on articulation scores noted. Comparison of foreign accent with other types of speech distortion, and analysis of their interactions, has lead to a distinction between signal-dependent and signal-independent distortion and their effects in degrading speech. Initial findings on the relations among ratings of foreign accent, amount of English training, masking, and intelligibility are presented in Appendix A. Research is in progress to compare the effects of filtering native and foreign speech with the effects of speech distortion already noted.

2. The effects of schedules of reinforcement on properties of the vocal response correlated with rate.

Emission of the relatively simple vocal response /u/ was reinforced with points under the following sequence of schedules: continuous reinforcement, variable-interval reinforcement, extinction. Tape recordings of the experiment were processed to determine the relative amplitude, pitch, and duration of each response and these data were correlated with cumulative records of the rate of responding. Initial findings are presented in Appendix A. Research is in progress to examine the relations among the several properties of the vocal response with other subjects and other schedules of reinforcement. Because instrumentation is available for the precise specification of several properties of the vocal response, this response is an excellent vehicle for the study of certain basic problems in conditioning.

3. Self-shaping of vocal behavior.

Several Thai tonemes, rendered by a linguist, were tape recorded individually on magnetic tape loops and presented repeatedly to American undergraduates. The subject was instructed to imitate the sound between presentations and to continue practice until he generated "a completely faithful reproduction." S was then trained to discriminate among the sounds of Thai and the self-shaping process repeated. Tape recordings of the experiment were then analyzed to permit comparison of the model pattern with the subjects' matching behavior before and after discrimination training. Initial findings are presented in Appendix A. Research in progress is aimed at assessing the effects of various kinds of discrimination training on the efficacy of self-shaping and at extending findings with the tonemes to other segmental and suprasegmental features of speech.

4. Equal loudness contours.

The theoretical and applied importance of the equal loudness contours seemed to warrant their further investigation. Advances in acoustic and psychophysical measurement permit a continuous determination of the form of these contours. An oscillator slowly scans the audio-frequency range while the signal transduced by the headphones is subtracted from the input waveform. When the resultant amplitude-modulated sweep-frequency signal is tape recorded and played back to the same headphones, their frequency response is effectively flat. The subject is instructed to adjust a sone potentiometer so as to maintain the signal at constant loudness despite changes in pitch, while a graphic level recorder gives a continuous record of the amount of attentuation introduced by S. This process is, of course, repeated at several intensity levels. With problems of instrumentation nearly solved, the experiment is about to begin.

5. The effects of changing vowel parameters on perceived loudness and stress. IV: The reception and production of vowel duration.

This study is the fourth in a series which seeks to specify the parameters of stress perception. As in Experiments I, II, and III (see Research Completed), ratio-scaling techniques are employed to develop subjective scales that permit a prediction of stress estimation and matching. The first three studies were devoted to an analysis of the variables contributing to vowel loudness, a major parameter of linguistic stress. This study investigates the reception and production of duration. Initial findings indicate that subjective scales for duration, unlike loudness, are nearly linear against their physical correlate, vowel duration.

6. The effects of changing vowel parameters on perceived loudness and stress. V: Predicting the acoustic parameters of linguistic stress.

Continuing the analysis of stress developed in Experiments I - IV, this study first charts the subjective scales for received and produced vowel pitch. Initial findings indicate that the speaker's scale of his own vocal pitch and the listener's scale of vowel pitch are both nearly linear against the fundamental frequency of the vowel. With a knowledge of the sensory dynamics of received and produced loudness, duration, and pitch and their interactions, we have encompassed the acoustic dimensions of stress. It is then possible to make quantitative predictions concerning the perception and matching of linguistic stress in a natural language. Research in progress is aimed at assessing the accuracy of these predictions.

7. Shaping the prosodic features of speech with an auto-instructional device.

If vocal behavior is to be shaped, the experimenter must specify the dimensions of responding to be altered and the terminal behaviors desired. If the shaping is to be effected by an auto-instructional device, this device must be capable of (1) analyzing the relevant response dimensions in real time, (2) evaluating the response with respect to the desired terminal performance, (3) adjusting reinforcement contingencies as a function of the behavior of the subject, (4) providing reinforcement.

The feasibility of such a device is greatly enchanced if the number and complexity of response dimensions to be shaped is limited; this compromise may also facilitate analysis of the conditioning process. A device has been designed to shape these prosodic features of speech: fundamental frequency, relative amplitude, and tempo. The segmental features of speech are not treated. The name of the device is SAID (speech auto-instructional device).

SAID can perform in any one or more of three "modes": pitch, amplitude, and tempo. Auditory stimuli (speech or non-speech) are recorded and played back by the device. The student is instructed to respond echoically either (a) concurrently or (b) in alternation with the stimulus sequence.

SAID analyzes the selected dimensions of the stimulus and echoic response, compares them, and generates an error signal proportional to the difference as a function of time. This error voltage is available to the experimenter for graphic recording and to the subject, if the experimenter chooses. Alternatively, the subject may view a discrete signal at the completion of his echoic chain that indicates whether the total error voltage in the prescribed mode(s) is less than an error threshold selected by the experimenter. This error threshold may be varied from trial to trial.

A device for shaping the prosodic features of speech should prove valuable not only for basic research in the control of vocal behavior but also for such applied problems as second-language learning and aphasic reconditioning.

8. On the relations between stimulus generalization and psychophysics.

In this study an attempt is made to relate stimulus generalization to psychophysical scaling by obtaining magnitude estimations of vowel loudness under two experimental conditions: (a) following discrimination training on five synthesized vowel sounds (/i/, /I/, /ɛ/, /æ/, and /a/) in which the vocal responses "ten" to the middle stimulus /ɛ/ is reinforced, and (b) following exposure to the same vowel sounds in which differential responding is not reinforced. The extent to which the effects of reinforcement generalize to other vowel intensities not present during training may be reflected in differences in shape and slope of the functions relating magnitude estimation to auditory intensity. These findings may be compared to those obtained under a third condition in which vowel generalization along the intensity continuum is observed without magnitude estimation instructions.

APPENDIX A. RESEARCH IN PROGRESS. INITIAL FINDINGS

SELF-SHAPING OF VOCAL BEHAVIOR

PROPERTIES OF THE VOCAL RESPONSE CORRELATED WITH RATE OF EMISSION

FOREIGN ACCENT AND SPEECH DISTORTION

SELF-SHAPING OF VOCAL BEHAVIOR

Appendix A: Initial Findings

Self-shaping of Vocal Behavior

In one stage of second-language learning, the student is required to imitate certain foreign utterances, that is, to match his vocal response to several properties of a complex speech signal. This echoic behavior has a sensory (discriminative) and a motor (differentiative) component. If an experimenter wishes to condition this behavior, he may, at first, reinforce any response emitted in the presence of the discriminative stimulus, and then differentially reinforce (shape) successive approximations to the terminal behavior that is desired. The classroom language learning situation, however, is different: the student, not the experimenter, decides what is to be considered an approximation to the desired response, and which approximations are to be reinforced.

This shift in behavioral control reveals an implicit assumption concerning the training of the experimenter. Certain auditory descriminations are required if vocal behavior is to be shaped. (It is, of course, impossible to differentially reinforce successive approximations to a terminal response unless these approximations are discriminated.) The experimenter or teacher is usually trained in these discriminations; usually, the student is not. As a consequence, the *self*-shaping process may change the topography of the vocal response without leading to more accurate echoic behavior. If the subject were first trained to discriminate the relevant properties of the speech signal he might be more likely to discriminate changes in his own vocal behavior and to progress toward a more accurate echoic response.

Two experiments were performed to assess the effects of discrimination training on the self-shaping of echoic behavior. In Experiment I each subject echoed the Thai toneme, /kā/, repeatedly until he believed he gave a completely accurate reproduction. In Experiment II, the above procedure was employed twice, once before and once after discrimination training.

Experiment I

Method

Each of three undergraduates served individually in sessions lasting approximately 15 minutes. S was seated in an anechoic chamber in front of a cartridge tape deck and microphone. He wore a binaural headset with high fidelity earphones (PDR-8) mounted in doughnut cushions (MX-AR/41), which attenuated air-conducted side tone by about 15 db. The discriminative stimulus, /ka/, rendered by a linguist, was recorded on a loop of magnetic tape and presented repeatedly at 1.6 second intervals through one earphone. An amount of sidetone was introduced to the other earphone which approximately compensated for the attenuation introduced by the headset. The following instructions were read to the subject.

> "In front of you is a cartridge containing magnetic tape. When you place the cartridge on the tape deck, like this, you will hear a sound repeated rapidly. Your task is to imitate the sound as accurately as possible. Continue to listen to the sound and imitate it between presentations, until you believe you have given a completely faithful reproduction. Remember, and I wish to emphasize this point, your task is to reproduce the sound exactly. Since you are being paid according to how long you work, it will be to your advantage to repeat the sound until you have faithfully reproduced it."

In order to measure the duration of the discriminative stimulus and the subject's responses, tape recordings were processed in this manner: the speech signals were sent to an average speech power circuit whose d-c output triggered an interval timer (Hewlett Packard 522 frequency counter). The duration of each signal, in milliseconds, was then recorded by a print out counter. The fundamental frequency was selected from the complex speech signal by band-pass filtering (100-150 cps.) and then sent to the frequency counter and associated printout. A pitch slope was computed by using the above circuitry to print out the fundamental frequency at 175 msec. intervals, beginning with the onset of the signal. Frequency change, in cps./msec. (pitch slope), was then given by

$$\frac{F_n - F_o}{175(N-1)}$$

F_o is the initial pitch
F_n is the terminal pitch
N is the number of readings.

The discriminative stimulus /kā/ had a duration of 600 msec., a terminal pitch of 125 cps., and a relatively flat pitch slope of -.013 cps./msec.

Results and Discussion

Figure 1 summarizes changes in response duration and pitch slope during self-shaping by each of three subjects. The mean response duration of S_1 shows a slight increase over the session. This parameter for S_2 is more variable, and increases in length toward the end of self-shaping. For S_3, response duration falls at the beginning of the session and then shows a slight increase. In all three cases, response duration changes slightly during the session but there is no evidence for a trend in the direction of approximating the duration of the discriminative stimulus (S^D).

The mean pitch slope of the vocal responses emitted by S_1 falls during the first part of the session and then stabilizes at an extremely steep value, one that is far steeper than that of the S^D. The pitch slope for S_2 diverges slightly from that of the S^D before stabilizing. The responses of S_3 are characterized by a rising intonation. The plot of the mean pitch slope increases at first, and then stabilizes.

The value of the pitch slope for two of the three subjects does not approximate the S^D. Furthermore, the mean pitch slope of the three subjects does not approach any common value.

The standard deviation about the mean pitch slope of blocks of consecutive responses shows no evidence for a systematic decrease in the variability of responding (see Fig. 1).

Experiment II

Method

In this experiment, discrimination training was interpolated between two replications of Experiment I. These instructions were read to each of three subjects.

> "During the second phase of this experiment you will hear a series of stimuli. You are to pull this lever when you hear the first sound in this series and every time afterward that you hear the same sound. When you do this you will accumulate points on this counter. If you

respond to the wrong stimulus you will lose points on the counter. We want you to try to accumulate as many points as you can. Remember, you accumulate points by pulling the lever to the first sound in the series and by pulling the lever every time that the sound appears. If you pull the lever to the wrong sound you will lose a point. You will be expected to reach a certain criterion score on these series of sounds. If you do not reach criterion the first time, your counter will be cleared and after a slight delay we will begin over."

A tape recording of five Thai tonemes, /kā/, /ká/, /kà/, /kâ/, and /kǎ/ was presented to the subject. Each toneme appeared eight times in irregular order at four-second intervals. As in Experiment I, the positive discriminative stimulus was /kā/. The operant response was a pull on a Lindsley manipulandum, which was reinforced in the presence of S^D by an increase in the number displayed on a glow tube.

Results and Discussion

Figure 2 shows the mean response duration of blocks of consecutive responses for the three subjects before and after discrimination training. Response duration for S_4 shows considerable variability both in the pretest and postest. The mean duration rises toward the end of the pretest. In the postest the mean duration falls although it always exceeds the corresponding mean observed during pretest. The duration data for S_5 are similar: during pretest, response duration is less than during postest. As was observed for S_4 and S_5, the effect of discrimination training on S_6 is, in general, an increase in response duration.

The pitch slope data of the three subjects, shown in Figures 3 and 4, are similar to those observed in Experiment I. There is a systematic trend toward a terminal performance. This terminal behavior differs among subjects, and it does not approximate the acoustic parameters of the S^D in both the pretest and the postest. The effect of discrimination training is to flatten the pitch slope of the responses by two of the three subjects; the reverse effect is observed for a third subject, however.

The standard deviation about the mean pitch slope of blocks of consecutive responses shows no evidence for a systematic decrease in the variability of responding during self-shaping. Furthermore, Figs. 3 and 4 show that discrimination training has little effect on this parameter.

Summary

Six subjects served in two experiments to assess the effects of self-shaping and discrimination training on the topography of an echoic vocal response.

(1) During "self shaping," the duration and pitch slope of echoic vocal responses tend to stabilize at some value.

(2) This "steady state" does not necessarily have the same acoustic parameters as that of the discriminative stimulus.

(3) Discrimination training tends to reduce the overall departure of the topography of echoic responding from that of the discriminative stimulus.

Figures

Fig. 1. Bottom: mean response duration (msec.) of blocks of consecutive responses for S_1, S_2, and S_3. Top: mean pitch slope (frequency change per msec.), and the standard deviation about the mean, for blocks of consecutive responses. Number of responses per block: S_1, 5, except 4 for block 7; S_2, 10, except 6 for block 6; S_3, 25.

Fig. 2. Mean response duration during pretest and postest (msec.) of blocks of consecutive responses for S_4, S_5, and S_6. Number of responses per block: S_4, 20, except 8 for block 7 (pretest) and 18 for block 4 (postest); S_5, 20, except 13 for block 11 (pretest) and 6 for block 8 (postest); S_6, 5.

Fig. 3. Mean pitch slope (frequency change per msec.) during pretest (unshaded symbols) and postest (shaded symbols), and the standard deviation about the mean for blocks of consecutive responses by S_6 and S_4. The number of responses in each block are the same as in Fig. 2.

Fig. 4. Mean pitch slope (frequency change per msec.) during pretest (unshaded symbols) and postest (shaded symbols), and the standard deviation about the mean for blocks of consecutive responses by S_5. The number of responses in each block are the same as in Fig. 2.

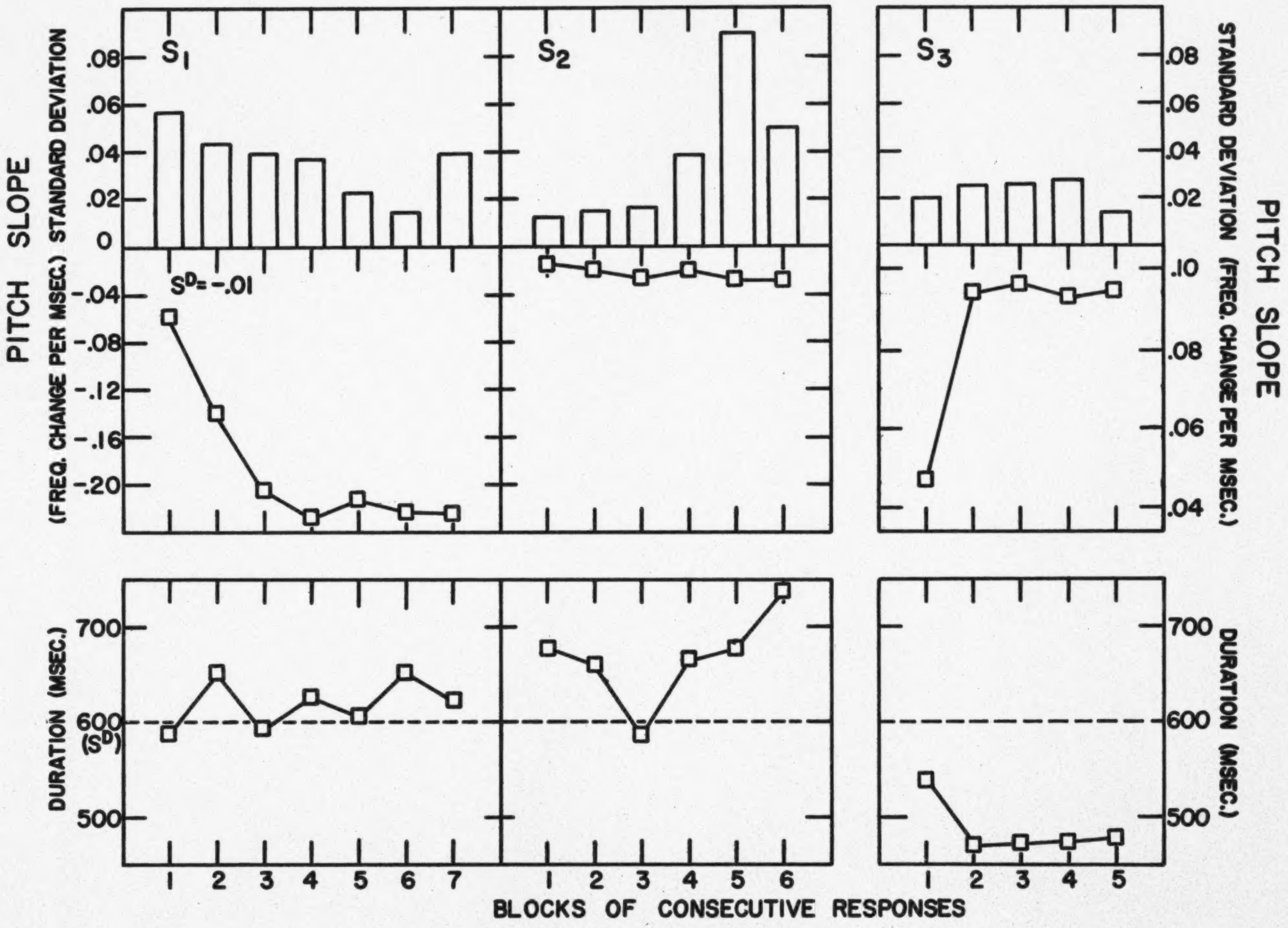
PITCH SLOPE
STANDARD DEVIATION
(FREQ. CHANGE PER MSEC.)
S1
S2
S3
S^D = -.01
.08
.06
.04
.02
0
-.04
-.08
-.12
-.16
-.20
.10
.08
.06
.04
DURATION (MSEC.)
700
600 (S^D)
500
1
2
3
4
5
6
7
BLOCKS OF CONSECUTIVE RESPONSES

Fig. 1

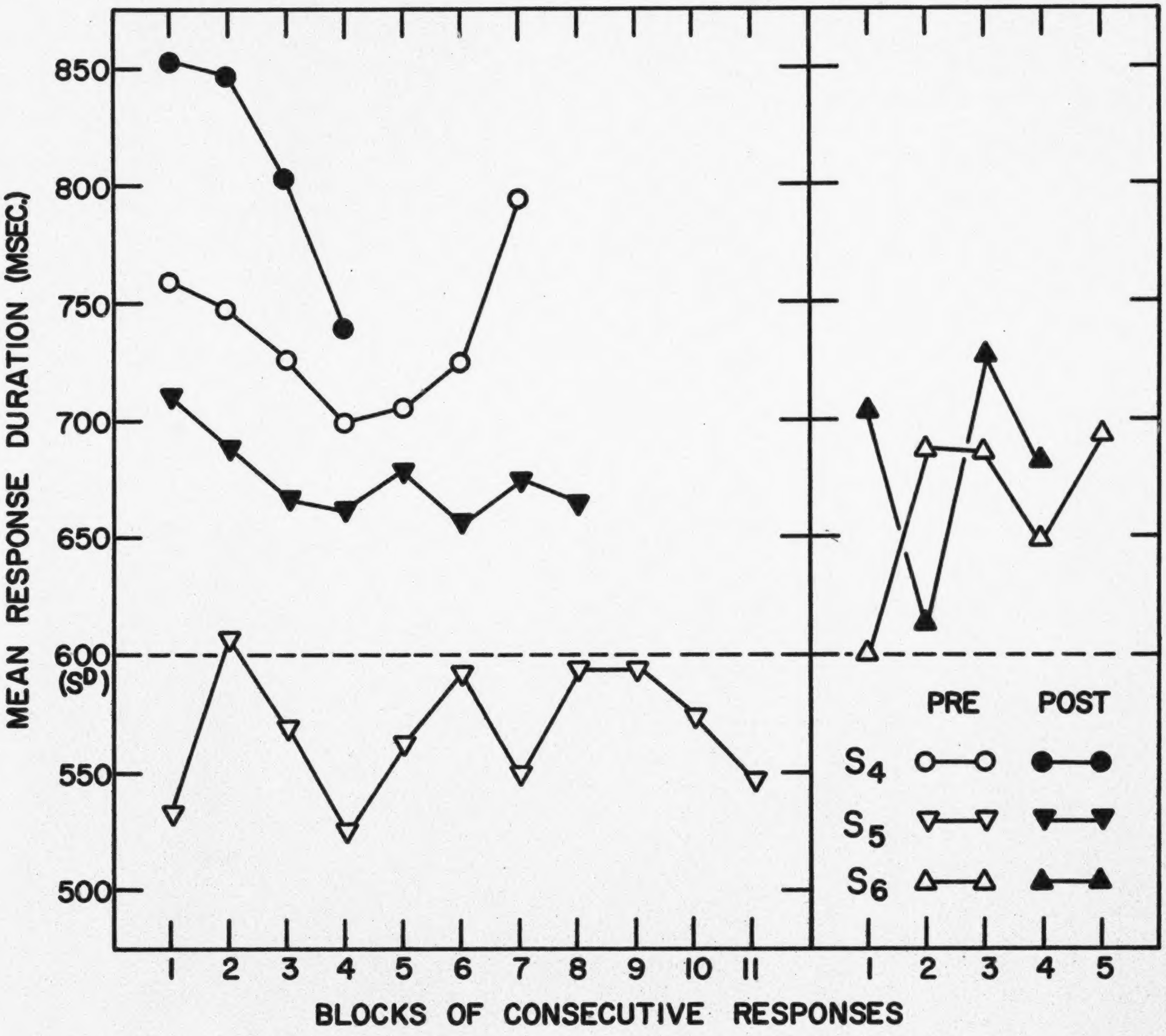
MEAN RESPONSE DURATION (MSEC.)
850
800
750
700
650
600
(S^D)
550
500
1
2
3
4
5
6
7
8
9
10
11
1
2
3
4
5
BLOCKS OF CONSECUTIVE RESPONSES
PRE
POST
S_4
S_5
S_6

Fig. 2

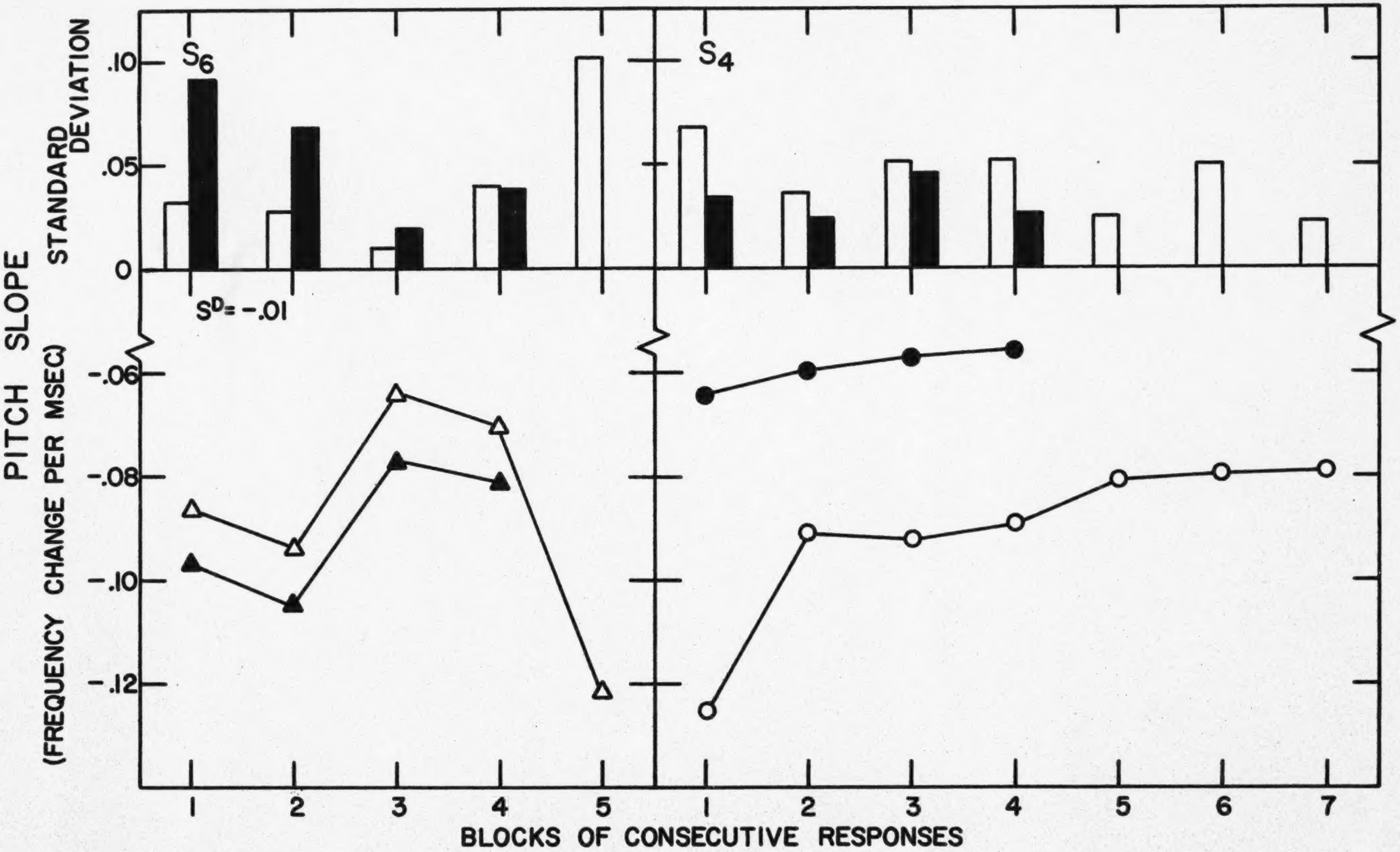
S6
S4
STANDARD DEVIATION
.10
.05
0
SD= -.01
PITCH SLOPE
(FREQUENCY CHANGE PER MSEC)
-.06
-.08
-.10
-.12
1
2
3
4
5
1
2
3
4
5
6
7
BLOCKS OF CONSECUTIVE RESPONSES

Fig. 3

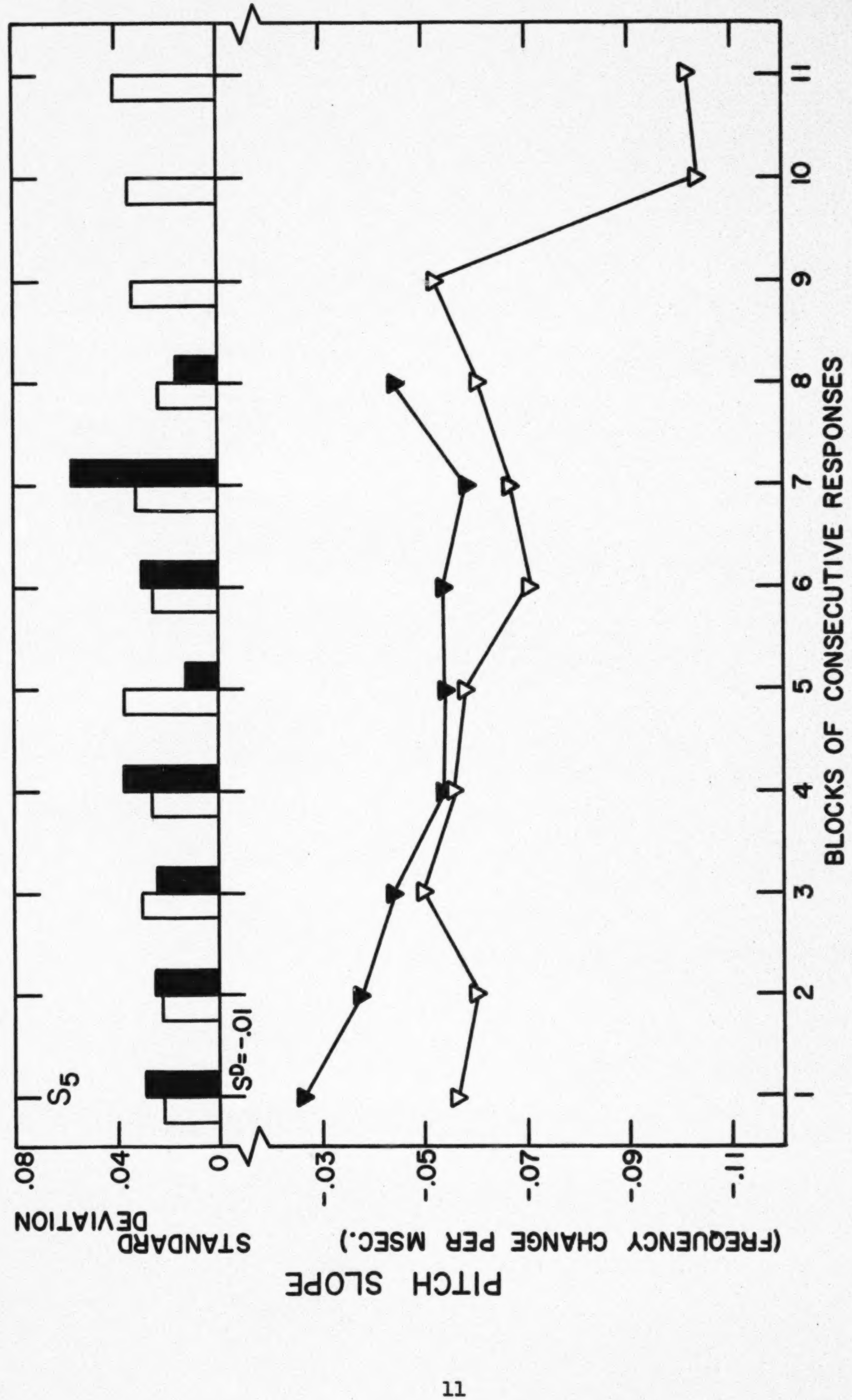
PITCH SLOPE
(FREQUENCY CHANGE PER MSEC.)
STANDARD DEVIATION
.08
.04
0
-.03
-.05
-.07
-.09
-.11
S5
SD=-.01
1
2
3
4
5
6
7
8
9
10
11
BLOCKS OF CONSECUTIVE RESPONSES

Fig. 4

PROPERTIES OF THE VOCAL RESPONSE CORRELATED WITH RATE OF EMISSION

Appendix A: Initial Findings

Properties of the Vocal Response Correlated with Rate of Emission

Several studies have shown that human and subhuman vocalizing are amenable to operant control. Therefore, the vocal response may be used as a vehicle for the investigation of certain basic problems in conditioning. It is often the response of choice because instrumentation for the analysis of speech is well advanced; changes in response topography, duration, and amplitude, as well as rate of emission, may be measured with great facility and accuracy.

The present study describes the changes that take place in the pitch, average speech power, duration, and rate of emission of a vocal response (the phoneme /u/) under a sequence of three schedules of reinforcement: continuous reinforcement (crf), variable-interval reinforcement (VI), and extinction (ext). (For an account of the effects of schedules of reinforcement on the rate of responding, see Ferster and Skinner, 1957.)

Method

A female undergraduate, aged 20, was seated in an anechoic chamber in front of a microphone and loudspeaker. Her head was taped with adhesive to a headrest to maintain a constant distance between subject and microphone. The instructions to the subject also summarize the procedure:

> "This is an experiment in speech. You will hear numbers read to you over the loudspeaker in groups of about five or six. Each time a group of numbers is read, your job is to write down the numbers in a row of cells on your response sheet. Start a new row for every group of numbers. Numbers are presented only when you say /u/ into the microphone in front of you. Try not to make any other sounds at all, as this may disturb the experiment. The object is to see how many numbers you are able to write down correctly during the experiment, which will last about two hours. Try and stay in the position the experimenter puts you in, throughout the experiment. Are there any questions? The experiment will begin a few seconds after I leave the room."

Each of the first twenty responses was reinforced by a four-second tape recorded sequence of random numbers presented over the loudspeaker. Subsequently, vocal responding was reinforced on a variable interval schedule with a mean interval of 64 seconds. After 25 reinforcements on VI, (a total of 72 minutes) the loudspeaker was disconnected and extinction was in effect for 48 minutes.

Each vocal response closed a voice-operated relay which, in turn, operated a cumulative recorder, yielding a continuous record of the rate of responding. Tape recordings of the subject's vocal responses were processed electronically to measure the duration, pitch, and relative amplitude of each response. (1) Duration measurements were obtained by sending the recorded signal to an average speech power circuit (integrating time 10 msec.) whose output triggered an electronic counter (Hewlett-Packard 522B). The counter measured the duration in milliseconds and sent an analog voltage to a print out counter. (2) Pitch measurements were obtained by filtering the speech signal so as to select the fundamental frequency, converting the sinusoid to a d-c voltage of proportional amplitude (Hewlett-Packard frequency converter 500BR), and recording this voltage on a graphic level recorder. By means of a calibration, the height of each tracing was converted to cycles per second. (3) Amplitude measurements were obtained by sending the tape recorded signal to an average speech power circuit and recording the logarithm of the output voltage on a calibrated oscillograph (Minneapolis-Honeywell Visicorder). The height of each tracing was then converted to decibels with the peak speech power of the weakest response serving as a reference.

Results and Discussion

Figure 1 presents the cumulative record of responding. The deceleration in responding following crf, the low rate of responding sustained under VI, and the further deceleration during extinction are all characteristic of the effects of these schedules on other human and subhuman operants. Figure 2 plots the cumulative pitch and cumulative amplitude of successive responses. Both the pitch and amplitude of responding decrease during crf and increase rapidly during the extinction interval preceding the first VI reinforcement. These findings are similar to those obtained by Notterman (1959) in an investigation of the force of bar-press in the rat under crf and ext. During the remainder of the session, the amplitude of responding does not fluctuate appreciably. Lane (1960) has reported the same observation with the human vocal response /u/ under a drl 15 sec. schedule of reinforcement. The pitch of the vocal response, however, varies extensively. Following most, but not all, reinforcements, there is a local decrease in pitch.

Table I permits a comparison of the mean and standard deviation of the pitch, amplitude, and duration of vocal responses under the three experimental conditions. The average pitch, amplitude, and duration are all higher during VI conditioning than during crf and higher during extinction than VI. Variability in pitch and in duration is greater under extinction than VI. This finding resembles that reported by Antonitis (1950) who measured the variability of a nose insertion response in the rat. The average deviation around the median position of the animal's nose in a horizontal slot was higher during crf and extinction than during periodic reconditioning---essentially the findings of the present experiment.

The duration parameter of the vocal response showed the largest change (over 60 per cent) under the conditions of the present experiment. However, the duration, pitch, and amplitude of the vocal response were all found to be highly correlated. The product moment correlation coefficients, determined for vocal responses during VI, were: duration and pitch, 0.68; pitch and amplitude, 0.58; amplitude and duration, 0.63.

Summary

A human subject emitted the vocal response /u/ under three schedules of reinforcement: crf, VI, ext. Changes in the rate of responding were similar to those obtained with other human and subhuman operants under comparable schedules of reinforcement. The mean pitch, amplitude, and duration of responding were higher during VI than during crf and highest during extinction. Variability in these parameters of the vocal response was

higher during crf and extinction than during VI. All three dependent variables were highly correlated ($r > 0.58$).

References

Antonitis, J.J. Variability of response in the white rat during conditioning and succeeding extinction and reconditioning. Ph. D. dissertation, Columbia Univ., 1950.

Ferster, C.B. and Skinner, B.F. *Schedules of Reinforcement*. New York: Appleton-Century-Crofts, 1957.

Lane, H.L. Temporal and intensive properties of human vocal responding under a schedule of reinforcement. *J. exp. anal. Behav.*, 1960, 3, 183-192.

Notterman, J.M. Force emission during bar-pressing. *J. exp. Psychol.*, 1959, 58, 341-347.

Table I

The Mean and Standard Deviation of the Pitch, Amplitude, and Duration of Vocal Responses During Crf, VI, and Ext

Condition	Number of Responses	Pitch (cps.)		Amplitude (db)		Duration (msec.)	
		M	SD	M	SD	M	SD
Crf	20	212.	16.	21.	2.6	234.	48.
VI 64 sec.	130	226.	14.	26.	1.4	329.	32.
Ext	90	227.	18.	27.	1.4	382.	44.

Figures

Fig. 1. Cumulative record of vocal responses by one human subject. At *a* the schedule of reinforcement was changed from crf to VI 64 sec. and at *b* from VI to ext.

Fig. 2. Cumulative amplitude and cumulative pitch of vocal responses by one human subject under three schedules of reinforcement.

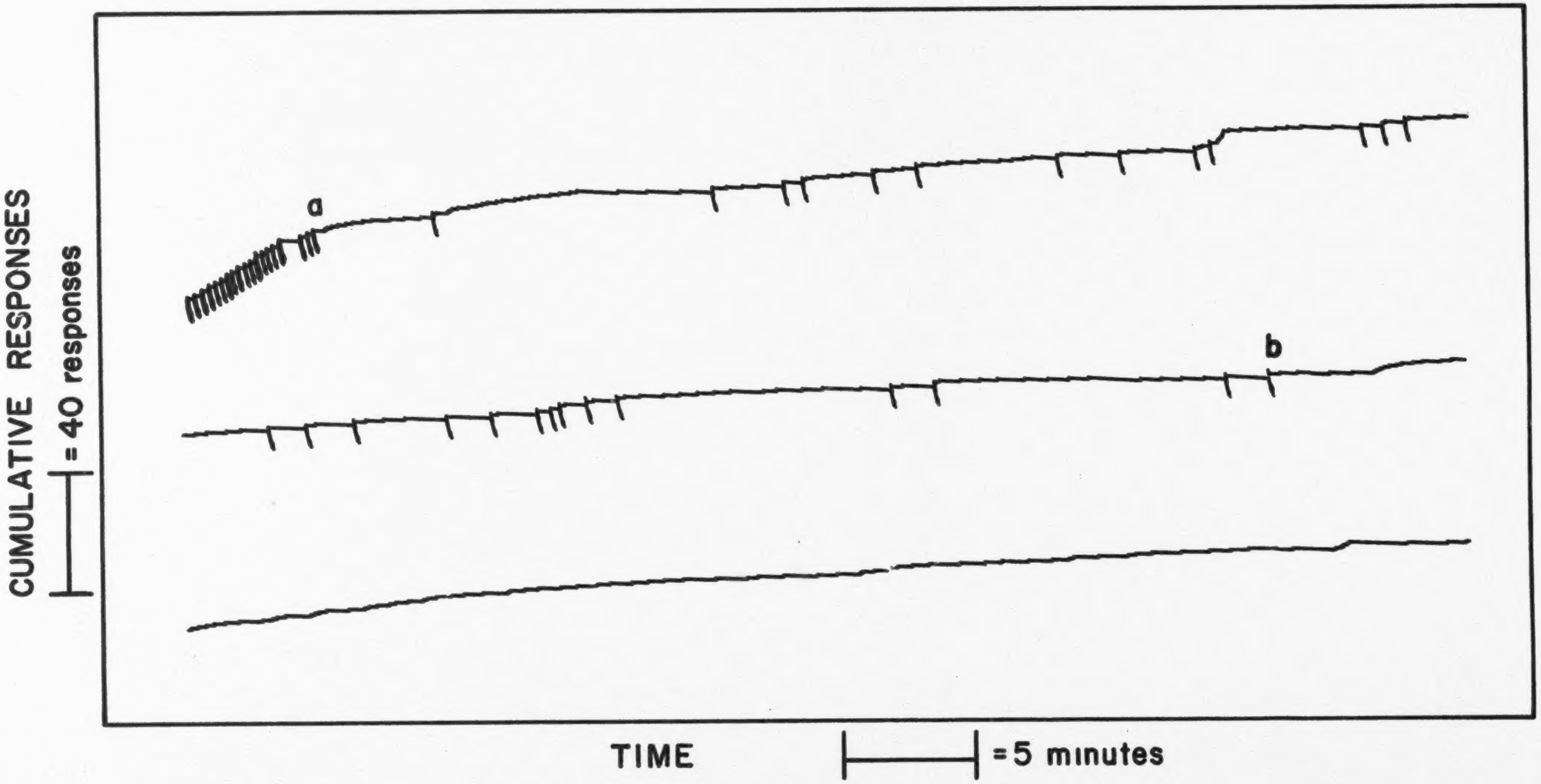

Fig. 1

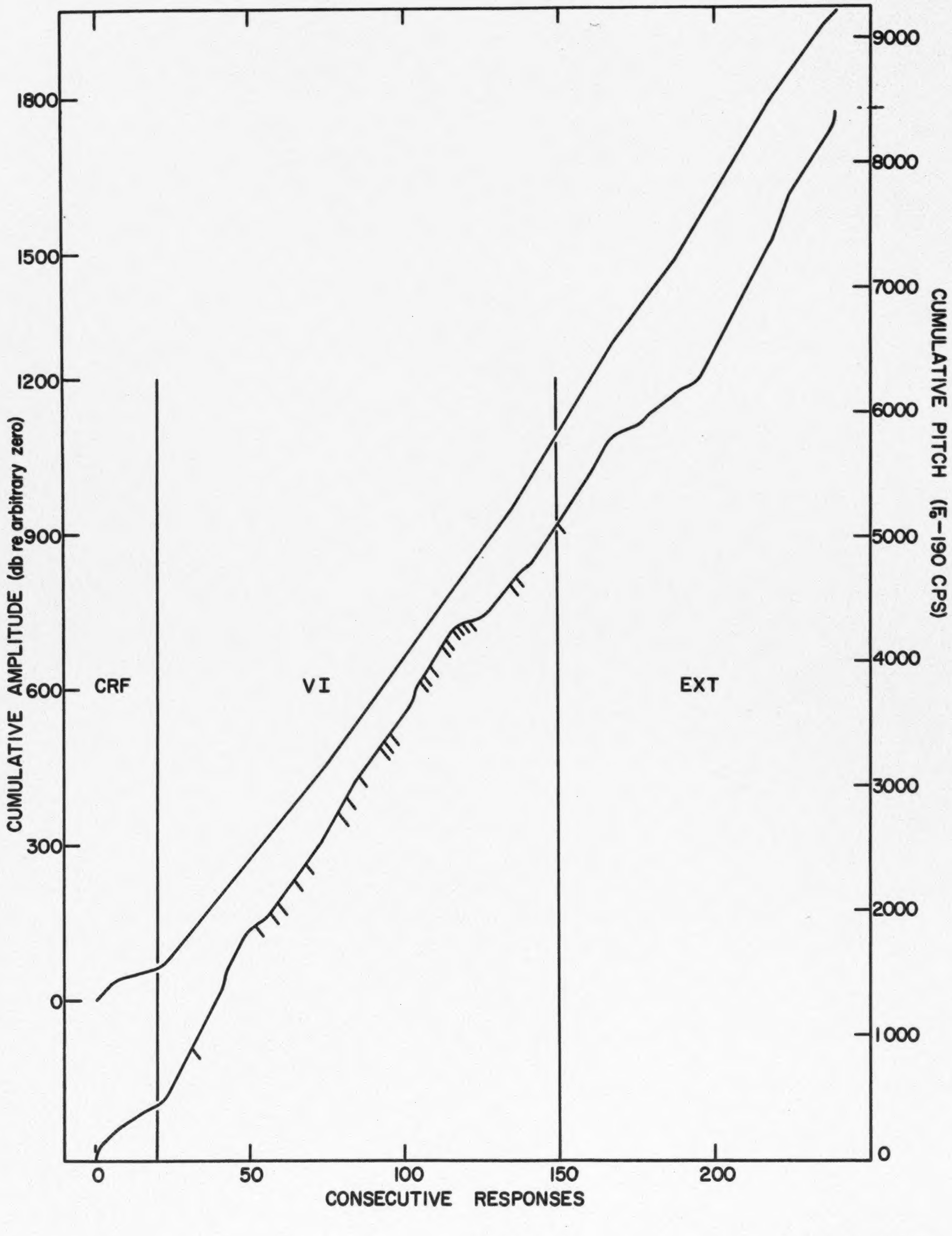
CUMULATIVE AMPLITUDE (db re arbitrary zero)
0
300
600
900
1200
1500
1800
CUMULATIVE PITCH ($\bar{F}_0$-190 CPS)
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
CRF
VI
EXT
0
50
100
150
200
CONSECUTIVE RESPONSES

Fig. 2

FOREIGN ACCENT AND SPEECH DISTORTION

Appendix A: Initial Findings

Foreign Accent and Speech Distortion

Among the methods that have been used to measure the effects of distortion on speech communication, articulation tests, subjective appraisals, and threshold tests are the most common. This study employs the first two methods to assess and compare the effects of foreign accent and masking noise on the intelligibility of speech.

Foreign accent may be considered a type of signal-dependent speech distortion; the nature and extent of the "accent" depends, in part, on the original signal to be rendered. Masking noise, however, is a type of signal-independent speech distortion since, typically, the spectrum and intensity of the noise are independent of the masked signal. Several decades of research have shown that speech perception is relatively unaffected by this latter kind of distortion (e.g., masking, filtering, time-sampling). Experimental manipulation of the time, frequency, or amplitude dimensions of speech must excede normally-encountered ranges of signal distortion by a wide margin before intelligibility is impaired appreciably.

Foreign accent, on the contrary, may effect a dramatic reduction in intelligibility. In this respect it is like other types of signal-dependent speech distortion, such as baby talk, dysarthric speech, and dialects. In a discussion of second-language learning, Liberman *et al.* (1957) have indicated some of the variables that may underlie the perception of speech distorted by foreign accent: "If [the listener's] discriminations have, by previous training, been sharpened or dulled according to the position of the phoneme boundaries of his native language, if the acoustic continua of the old language are categorized differently by the new one, then the learner might be expected to have difficulty perceiving the sounds of the new language until he has mastered some new discriminations, and perhaps, unlearned some old ones."

Similarly, the subject who is attempting to identify spoken words in his own language rendered with a foreign accent must also categorize stimuli from familiar acoustic continua in unfamiliar ways. For example, when listening to English rendered by a German with considerable foreign accent, the English listener must classify the acoustic complex /zi/ (appearing in English) as /θi/ if he is to achieve correct recognition. Other examples fill the repertoire of popular comedians and mimics.

Most of the prior research on intelligibility has been concerned with signal-independent speech distortion. The masking variable of the present study represents, therefore, a "standard" distorting operation whose effects may be compared to those of foreign accent. The experimental design also permits an assessment of the interaction effects of the two types of distortion operating in concert.

Method

The first measure of intelligibility employed in the present study is articulation score (see Egan, 1948). Typically, an announcer reads a set of syllables, words, or sentences to a group of listeners, and the percentage of items correctly recorded by those listeners is called the articulation score. Lists of English monosyllabic words were used in this study in preference to *nonsense syllables* to avoid artifacts introduced in training announcers to pronounce these sounds and to free the listeners from phonetic transcription. English *sentences* were avoided, as textual cues could affect the intelligibility of individual words.

Four "PB" lists of 50 words each were constructed from phonetically balanced sets compiled by the Harvard Psycho-Acoustic Laboratory. These sets attempt to provide items of monosyllabic structure, equal average difficulty of intelligibility, composition representative of English speech, and words in common usage. In the articulation scoring, transcriptions arising from homonymous forms of an item were considered correct.

The independent variable of foreign accent was instrumented by using four male speakers who spoke the following native languages: (a) English; (b) Serbian; (c) Punjabi; and (4) Japanese. Each speaker of a foreign native language was an undergraduate student at The University of Michigan and was obtained through the Foreign Language Institute there. It may be of interest to note The University of Michigan English Proficiency Test scores for the three foreign speakers: Serbian, 72; Punjabi, 87; Japanese, 80. A score below 90 is generally taken to indicate an inadequate command of English for successful University study.

Each of the four speakers read the four PB lists in a different order at the rate of one word every five seconds; 30 seconds were allowed to elapse between lists. The tape recorded articulation lists were then copied onto a second tape recorder and the record level adjusted so as to maintain a constant peak amplitude (10 db below 0 VU or approximately 50 db SPL with TDH-39 earphones).

The independent variable of masking was instrumented by mixing the tape recorded signals with equal excitation noise at one of four levels to give four signal-to-noise ratios: 15, 4, -1.5, -5 db. These were selected arbitrarily to give an anticipated articulation score of 100 per cent for no accent-low masking, on the one extreme, and better than 0 per cent for foreign accent-high masking at the other extreme.

Twelve Midwest-American undergraduates, none of whom was acquainted with the native language of the three foreign speakers, served as listeners in groups of three. Each group was presented with the 64 stimulus series (4 speakers x 4 lists x 4 S/N ratios) in a different, counterbalanced order, so that each listener never heard a speaker read the same list twice, nor was the same list ever heard twice at the same noise level. After articulation testing, each group was presented with a different series of four PB lists; each list contained ten words read by one of the four speakers in the absence of masking noise. The subjects were instructed to rate the foreign accent of each speaker on a scale of 1 to 5 ("very little" to "very much").

Results

Table I shows the mean articulation scores of the four speakers and Table II the analysis of variance for these scores. Inspection of these tables reveals, in the first place, that foreign accent had a marked effect on intelligibility. *A posteriori* comparison of the weighted mean intelligibility scores for the English versus the three foreign speakers combined (Scheffe's method) shows a difference that is significant at the .01 level. It is also clear that masking noise degraded the intelligibility of speech. The interaction effect of noise and foreign accent is small and not significant. A product-moment correlation of the twelve ratings of foreign accent and the twelve articulation scores obtained by each speaker gave r = .11 (not significant).

Summary and Conclusions

Four speakers, three with strong foreign accents and one native American, read phonetically balanced lists of English monosyllables. Articulation scores and subjective ratings of foreign accent were obtained from twelve Americans listening under four diverse levels of masking with "white" noise.

1. Foreign accents lowered the intelligibility of speech for listeners unfamiliar with those accents.

2. Intelligibility of speech decreased as the sound pressure level of the noise increased, or as the signal-to-noise ratio decreased.

3. Foreign accent and noise did not interact in their effect on intelligibility.

4. The American speaker had lower ratings of foreign accent and higher intelligibility than the foreign speakers combined, but ratings of accent and intelligibility were not correlated within the foreign-speaker group.

References

Egan, J. P. Articulation testing methods. Laryngoscope, 58, 1948 955-991.

Liberman, A. J., Harris, K. S., Hoffman, H. S., and Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. J. exp. Psychol., 54, 1957, 358-368.

TABLE I

MEAN ARTICULATION SCORES AND FOREIGN-ACCENT RATINGS OF FOUR SPEAKERS

S/N ratio (db)	English	Japanese	Serbian	Punjabi
15	96.2	66.7	57.0	60.2
4	78.3	30.7	35.0	30.5
-1.5	69.8	27.2	31.0	21.2
-5	44.3	14.8	14.2	10.0
Foreign Accent Rating (1-5)	1.1	4.1	3.1	2.9

TABLE II

ANALYSIS OF VARIANCE OF THE ARTICULATION SCORES OBTAINED FOR FOUR SPEAKERS FROM TWELVE LISTENERS

Source	Mean Square	df	F	p
Speakers	18,429.12	3	176.72	< .001
Masking	20,074.72	3	120.61	< .001
Interaction	166.43	9	1.59	n.s.
Residual	104.28	176		

APPENDIX B. SPECIAL ACTIVITIES

Appendix B

Special Activities

A) The Project Director has presented (or will present) the following addresses:

1. Some differences between first and second language learning. Foreign Language Institute, Ann Arbor, Michigan. April, 1961.
2. Role of reinforcement in the control of vocal behavior. Convention on the Exceptional Child, Detroit, Michigan. April, 1961.
3. Application of operant conditioning to the machine-teaching of languages. First conference in Language Programming, The University of Michigan, Ann Arbor, Michigan. April, 1961.
4. Foreign language learning. 32nd Annual Summer Education Conference, Ann Arbor, Michigan. June, 1961.
5. Teaching machines. Engineering Summer Conference, The University of Michigan, Ann Arbor, Michigan. July, 1961.
6. Techniques of operant conditioning applied to second language learning. (With F. R. Morton.) XIVth International Congress of Applied Psychology, Copenhagen, Denmark. August, 1961.
7. The effects of changing vowel parameters on estimates of loudness. To be read at the 62nd Meeting of the Acoustical Society of America, Cincinnati, Ohio. November, 1961.
8. Shaping the prosodic features of speech with an auto-instructional device. To be read at a symposium of the American Association for the Advancement of Science, entitled "Verbal Behavior: the experimental analysis and controlled alternation of its formal properties," Denver, Colorado. December, 1961.

B) The following laboratories have been visited in the period February 1, to September 1, 1961.

Haskins Laboratories, New York, New York.

Psychological Laboratories, Harvard University, Cambridge, Massachusetts.

Psychological Laboratories, Columbia University, New York, New York.

Department of Phonetics, University of London, London, England.

Institute for Psychological Research, Oxford University, Oxford, England.

Psychological Laboratories, University of Copenhagen, Copenhagen, Denmark.

Institute of Telegraphy and Telephony, Royal Institute of Technology, Stockholm, Sweden.